

УДК 311 : 159.9

Зимовин Алексей Иванович, аспирант кафедры общей психологии
Харьковского национального университета имени В. Н. Каразина

e-mail: zimovin@ukr.net

БАЙЕСОВСКИЙ ФАКТОР ПРОТИВ P-ЗНАЧЕНИЯ: ОЦЕНКА ПРАВДОПОДОБНОСТИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ПСИХОЛОГИИ

Аннотация

Рассматриваются недостатки р-значения как статистического инструмента: связь с гипотетическими данными, намерениями исследователя, зависимость от размера выборки, отсутствие статистических доказательств. Интерпретационным определениям р-значения противопоставляется точное математическое: р-значение как вероятность данных полученных в исследовании при условии истинности нулевой гипотезы. Представляется альтернативный показатель для оценки правдоподобности статистических гипотез – байесовский фактор – отношение правдоподобия. Излагается его математическое и описательное понимание. Байесовский фактор представляет собой отношение вероятности данных при условии истинности одной гипотезы к вероятности данных при условии истинности другой. Приводятся способы его интерпретации: простой математический и основанный на концепции доказательств. Также рассматривается формула перехода от байесовского фактора к оценке вероятности самой гипотезы при полученных данных. Преимущества данного коэффициента показаны на примере из исследовательской практики, в котором применение классического фреквентистского подхода приводило к неверным интерпретациям (отсутствия, наличия различий), хотя использование байесовского фактора показало недостаточность доказательств как для нулевой, так и для альтернативной гипотез.

Ключевые слова: байесовский фактор, р-значение, статистика, правдоподобность гипотез, фреквентистский подход, байесовский подход, методологический плюрализм.

Зімовін Олексій Іванович, аспірант кафедри загальної психології Харківського національного університету імені В.Н. Каразіна

БАЙЄСІВСЬКИЙ ФАКТОР ПРОТИ Р-ЗНАЧЕННЯ: ОЦІНКА ПРАВДОПОДІБНОСТІ СТАТИСТИЧНИХ ГІПОТЕЗ У ПСИХОЛОГІЇ

Анотація

Розглядаються недоліки р-значення як статистичного інструмента: зв'язок з гіпотетичними даними, намірами дослідника, залежність від розміру вибірки, відсутність статистичних доказів. Інтерпретаційним дефініціям р-значення протиставляється точне математичне: р-значення як вірогідність даних отриманих у дослідженні за умови істинності нульової гіпотези. Представляється альтернативний показник для оцінки правдоподібності статистичних гіпотез – байєсовський фактор – відношення правдоподібності. Викладається його математичне та описове розуміння. Байєсовський фактор являє собою відношення вірогідності даних за умови істинності однієї гіпотези до вірогідності даних за умови істинності іншої. Наводяться способи його інтерпретації: простий математичний і заснований на концепції доказів. Також розглядається формула переходу від байєсовського фактору до оцінки вірогідності самої гіпотези при отриманих даних. Переваги даного коефіцієнта показані на прикладі з дослідницької практики, в якому застосування класичного фреквентистського підходу призводило до невірних інтерпретацій (відсутності, наявності розбіжностей), хоча використання байєсовського фактора показало недостатність доказів як для нульової, так і для альтернативної гіпотез.

Ключові слова: байєсовський фактор, p -значення, статистика, правдоподібність гіпотез, фреквентистський підхід, байєсовський підхід, методологічний плюралізм.

Zimovin Oleksii, Postgraduate Student of the Department of General Psychology, V. N. Karazin Kharkiv National University

BAYES FACTOR VS. P-VALUE: EVALUATION OF STATISTICAL HYPOTHESES LIKELIHOOD IN PSYCHOLOGY

Summary

The disadvantages of p -value as a statistical tool are considered, they are: relation with hypothetical data and researcher's intentions, dependence on the sample size, the lack of statistical evidence. Interpretive definitions of p -value are opposed a rigorous mathematical definition: p -value as the probability of the data obtained in the study under the condition of the null hypothesis truth. Alternative measure to evaluate the statistical hypotheses likelihood – Bayes factor or likelihood ratio – is presented. Its mathematical and descriptive understanding is outlined. Bayes factor is the ratio of the probability of the data under the condition of the truth of one hypothesis to the probability of the data under the condition of the truth of other. The methods of its interpretation are provided: simple mathematical and based on the evidence conception. The formula of transition from Bayes factor to the assessment of the hypothesis probability under the condition of the data obtained in the study is also considered. The advantages of this factor are shown in the example of research practice in which the use of classical frequentist approach leads to wrong interpretations (absence, presence of differences), although the use of Bayes factor shows insufficient evidence for both the null hypothesis and the alternative hypothesis.

Keywords: Bayes factor, p -value, statistics, hypotheses likelihood, frequentist approach, Bayesian approach, methodological pluralism.

Постановка проблемы. Проблема парадигмы психологической науки, имеющая ряд дополнительных друг другу решений (кризис парадигмы, мультипарадигмальность, внепарадигмальность), претерпевает все больший крен в сторону методологического плюрализма – множественности мнений, позиций, гипотез исследователей, которые могут сосуществовать и находить подтверждения в рамках эмпирических исследований. В то же время классический статистический подход оценки вероятности гипотезы, основанный на определении p -значения и соотнесении его с критическими уровнями ($p < 0,05$ и $p < 0,01$), не только не позволяет накапливать правдоподобность гипотез, но и, с другой стороны, демонстрирует чрезвычайно низкую воспроизводимость результатов, полученных в одном исследовании (на одной выборке), в других случаях. Актуальным в современной методологической ситуации, не только для психологии, но и для других наук, использующих статистические выводы, становится использование другого, альтернативного показателя, позволяющего сделать заключение о правдоподобности гипотезы, таким показателем является байесовский фактор или *отношение правдоподобия*.

Анализ публикаций. Критика p -значения и рассмотрение байесовского фактора в качестве его альтернативы подкрепляется многими обще- и конкретнонаучными публикациями. Сразу отметим, что p -значение является инструментом фреквентистского подхода в статистике. Еще в 1994 г. (в статье с ярким названием «Земля круглая ($p < 0,05$)») Дж. Кохен написал: «После 4 десятилетий жесткой критики, ритуал тестирования значимости нулевой гипотезы – механического дихотомического решения вокруг священного критерия 0,05 – все равно сохраняется» [4, с. 997]. Особенно много говорят о проблемах p -значения в медицине, где принятие, к примеру, решения об эффективности метода лечения только на основе значимости p , может в очень скором времени приводить к плачевным результатам. Дж. Йоаннидис отмечал, что отсутствие подтверждения научных открытий является следствием удобной, но необоснованной стратегии принятия

решения только на основе формальной оценки p -значения [9, с. 696]. Для иллюстрации неубедительности использования p нередко приводится обобщение факторов риска возникновения рака, которые были показаны в различных работах. Исследователи отмечают, что, на уровне $p < 0,05$, факторами возникновения онкологического заболевания являются: электробритвы, переломы рук (у женщин), люминесцентные лампы, аллергии, разведение оленей, работа официанта, содержание домашних птиц, низкий рост, высокий рост, хот-доги, и конечно же, холодильники! [6]. Не может не радовать иронически настроенного критика вывод С. Каназавы – агрессивные мужчины имеют больше сыновей [11, с. 450]. Все вышесказанное побуждает научное сообщество реагировать, иногда радикально. В частности, журнал «Общая и прикладная социальная психология» (Basic and Applied Social Psychology) объявил, что не будет публиковать статьи, содержащие p -значения, поскольку их слишком часто используют для обоснования исследований низкого качества [15]. Что с p -значением не так? Об этом – ниже, в изложении основного материала.

Байесовский подход в статистике называют альтернативным классическому, фреквентистскому (частотному) [3, с. 5], хотя известная теорема Т. Байеса была опубликована еще в 1763 г. и, в общем, байесовский подход такой же классический, как и фреквентистский. В чем разница? Фреквентистский подход в статистике пытается установить вероятность только после того, как некоторое число испытаний было проведено. В нем считается, что до испытаний никакой вероятности истинности гипотезы не существует, отсутствует априорная вероятность, тем самым элиминируется субъективность и создается претензия на абсолютную объективность такой статистики. Но позвольте, насколько это соответствует реальности? В какой ситуации исследователь не имеет никаких предположений относительно правдоподобности собственной гипотезы, никаких теоретических посылок, никаких предшествующих эмпирических фактов, свидетельствующих в ее пользу? Вопрос риторический. Ратуя за возвращение субъекта в науку,

Э. Морен писал: «Разве можно принять, что познание основывается на исключении познающего, что мышление основывается на исключении мыслящего, что субъект исключается из процесса конструирования объекта...? ... Разве можно принять такую ночь, опустившуюся над знанием?» [2, с. 36]. В байесовском подходе учитываются как апостериорные, так и априорные вероятности справедливости гипотезы. Более того, если основной задачей фреквентистского подхода становится подтверждение гипотезы (на самом деле, опровержение нулевой) с помощью получения данных (проведения испытаний), что характерно для дедукции, в байесовском подходе важен вопрос – «какая истина наиболее вероятна на основании существующих данных» [7, с. 997], то есть логика индуктивна, при этом учитывается вероятность не только одной гипотезы. Можно отметить возрастание количества публикаций, посвященных байесовскому подходу и в частности использованию байесовского фактора в исследовательской практике [8, 10, 12, 13, 14].

Цель данной статьи – рассмотрение сущности байесовского фактора и способов его интерпретации в психологических исследованиях. Так как отношение правдоподобия является альтернативой классическому p -значению, необходимо начать с более подробного освещения недостатков последнего.

Изложение основного материала.

Недостатки p -значения. Для того чтобы понять причины, по которым p -значение как статистический инструмент встречает все большую критику, а иногда и отвержение со стороны международного научного сообщества, необходимо понять саму его сущность и те заблуждения, которые имеются в его отношении.

Так, С. Гудман [8] сформулировал заблуждения в отношении p -значения, которые достаточно часто тиражируются в научных и учебных публикациях. Самым пагубным он считает заблуждение, которое можно передать следующим образом: если $p < 0,05$, то мы имеем только 5% шансов

ошибочного отклонения нулевой гипотезы (об отсутствии различий, корреляций), еще говорят о 5% вероятности допущения ошибки первого рода (ложноположительного результата). Почему это не так? Ответ кроется в самом устройстве p -значения: оно рассчитывается в предположении, что нулевая гипотеза истинна, и никак не может быть одновременно вероятностью ее неистинности. Среди последующих заблуждений можно назвать: убеждение, что при $p > 0,05$ различий между группами нет (на самом деле, это говорит лишь о том, что отсутствие различий лучше согласуется с результатами); при $p < 0,05$ статистически значимый вывод имеет клиническое значение, мы будем читать «значение для психологической практики» (но p не несет никакой информации о выраженности эффекта!); исследования с тем же p -значением обеспечивают те же доказательства против нулевой гипотезы (различные эффекты могут иметь одно и то же p -значение). Мы не будем называть все двенадцать рассмотренных Гудманом заблуждений, читатель может найти их в его работе, на которую мы сослались. Остановимся лишь на последнем заблуждении, которое на новом уровне суммирует все, что было сказано: научный вывод должен быть основан только на том является p значимым или нет. Тем самым с субъекта познания снимается ответственность, а сам он удаляется в изгнание. А вместе с ним внимание к величине эффекта (различий, корреляции), его предыдущие исследования, теоретические позиции, то есть все, что оказывало влияние, и было собрано в понятии априорной вероятности.

Откуда эти заблуждения? Возможно с одной стороны из попытки более простого определения p , или с другой – из желания найти хоть какое-то математически непогрешимое основание для решения сложных, системных научных проблем.

Что же такое, по сути, p -значение (интерпретируемое то как вероятность истинности нулевой гипотезы, то как вероятность ошибки, то еще в каких-то очень соблазнительных, но математически неточных терминах)? Со времен Фишера, который предложил использовать p -значение, и до сего дня, как вы

понимаете суть статистического инструмента не могла измениться, p -значение определяется как вероятность, при условии допущения отсутствия эффекта или отсутствии различий (нулевая гипотеза), получения равного, или более сильно выраженного результата, по сравнению с результатом, который наблюдался в данном исследовании [7, с. 997]. Отметим: p вообще не является вероятностью правдоподобности гипотезы – $P(H_0)$, или вероятностью правдоподобности гипотезы при таких полученных данных – $P(H_0/D)$! Наоборот, это есть вероятность получить такие же (или еще более различающиеся по группам, коррелирующие, то есть с более сильным эффектом) данные (если говорить корректнее, не сами данные, но критерий, который характеризует эффект: t , χ^2 , Z и т. д.), при условии допущения нулевой гипотезы – $P(D/H_0)$. Как и почему это так? Развернутое обсуждение сути и способа определения p выходит за рамки данной работы. Отметим лишь, что технически любое распределение вероятностей статистического критерия является нормальным. Соответственно вероятность получить средние значения статистики (t , χ^2 , Z и т. д.) является большой, но по мере увеличения выраженности критерия вероятность снижается, доходя, наконец, до точки 0,05, за которой обычно начинается светлое будущее современных научных открытий. То есть именно на распределении вероятностей (частот встречаемости, подход ведь фреквентистский) статистического критерия и определяется p -значение. Никакой магии и никаких отсылок к пресловутым вероятностям ошибок, гипотез и даже больше – существенности различий и т. п. Значение p просто показывает нам, насколько редко встречается то или иное значение критерия (t , χ^2 , Z и т. д.), если провести множество испытаний.

Э. Ярош и Дж. Уайли [10, с. 2-3] подытожили основные недостатки p -значения в четырех пунктах. Во-первых, p -значение связано с гипотетическими данными. Теоретическое распределение вероятностей критерия, с которым в итоге мы сравниваем эмпирический результат, на самом деле – именно теоретическое, оно – результат многократных итераций одного и того же испытания при условии правдоподобности нулевой

гипотезы, то есть не имеет никакого или имеет очень малое (здесь мы смягчаем названных авторов) отношение к реальным данным. Во-вторых, p -значение зависит от намерений исследователя. В подходе, на котором основывается p -значение, формирование выборки остается полностью на совести исследователя и лишь он отмечает момент, когда она сформирована. В-третьих, p -значение не предоставляет статистических доказательств. Для нас это один из самых важных пунктов. Этот недостаток легко понять, если обратиться к его разновидности – p -значение очень сильно зависит от размера выборки. И тогда, $p < 0,05$, для выборки в 30 человек и для выборки в 200, не могут предоставлять одинаковое количество доказательств против нулевой гипотезы. Выходит, что большая выборка, с крошечными эффектами (различия, корреляции), тем не менее, порождает малые p , то есть те, поиском которых классический исследователь и занят, но есть ли в этих «статистически значимых» эффектах хоть какая-то практическая польза? И последний недостаток, еще более важный, чем предыдущий – фреквентистский подход учитывает лишь экстремальность данных при нулевой гипотезе, без рассмотрения альтернативной. Как вообще можно делать заключение об истинности альтернативной гипотезы, просто на основании отклонения нулевой, не контролируя вероятность данных при условии истинности первой? На самом деле, вполне возможна ситуация, когда обе гипотезы должны быть отклонены, поскольку ни одна из них не соответствует полученным данным.

Сравнивая байесовский подход с фреквентистским, Э. Ярош и Дж. Уайли выводят преимущества первого: байесовский подход учитывает только наблюдаемые данные, момент сформированности выборки не является критическим, поскольку анализ данных одной выборки может быть использован в качестве информации для анализа следующей выборки, одинаковые байесовские факторы указывают на одинаковое количество доказательств в пользу альтернативной гипотезы, независимо от размера

выборки и других внешних факторов, байесовский фактор учитывает правдоподобности как нулевой, так и альтернативной гипотез.

Сущность байесовского фактора. Теорема Байеса знакома каждому, кто, изучая теорию вероятности, интересовался проблемами *условной* вероятности. Эта теорема имеет следующий вид:

$$P(H|D) = \frac{P(D|H) P(H)}{P(D)}, \quad (1)$$

где: $P(H|D)$ – апостериорная вероятность гипотезы при полученных данных, $P(D|H)$ – вероятность данных при условии истинности гипотезы, $P(H)$ – априорная вероятность гипотезы, $P(D)$ – полная вероятность данных (независимо от конкретной гипотезы).

В математической статистике для вычисления апостериорной вероятности нулевых и альтернативных гипотез в их отношении друг к другу используют такую формулу, по Э.-Я. Вагенмакерсу [16, с.790]:

$$\frac{P(H_0|D)}{P(H_1|D)} = \frac{\frac{P(D|H_0)P(H_0)}{P(D)}}{\frac{P(D|H_1)P(H_1)}{P(D)}} \quad (2)$$

Уравнение (2) упрощают, избавляясь от $P(D)$ общего для верхней и нижней частей. Тогда получается:

$$\frac{P(H_0|D)}{P(H_1|D)} = \frac{P(D|H_0) P(H_0)}{P(D|H_1) P(H_1)} \quad (3)$$

Это уравнение легко записать в текстовой форме:

$$\text{Апостериорная вероятность} = \text{Байесовский фактор} \times \text{Априорная вероятность} \quad (4)$$

С. Гудман еще называет апостериорную вероятность собственно вероятностью истины, априорную – субъективным компонентом, а байесовский фактор – компонентом доказательства [7, с. 997].

Из нашей математической и текстовой интерпретации становится понятной сущность байесовского фактора:

$$BF_{01} = \frac{P(D|H_0)}{P(D|H_1)} \quad (5)$$

Говоря строго, байесовский фактор – есть отношение вероятности данных при условии истинности нулевой гипотезы к вероятности данных при условии истинности альтернативной гипотезы. То есть байесовский фактор позволяет нам понять, в пользу какой гипотезы полученные данные свидетельствуют сильнее. Но об интерпретации ниже, сейчас следует заметить, что мы можем соотносить вероятности и иначе, получая другую форму байесовского фактора:

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} \quad (6)$$

Только от предпочтений исследователя зависит, выводить ли ему BF_{01} или BF_{10} . Иногда рекомендуют выводить BF_{10} тогда, когда цель – доказать альтернативную гипотезу. Возможно также вычисление логарифма из байесовского фактора, особенно это полезно при чрезвычайно больших BF . Этот логарифм называют *весом доказательств* [8, с. 1005].

Удачным практическим приложением байесовского фактора является собственно возможность вычисления апостериорной вероятности гипотезы [13]. Это та самая вероятность истинности гипотезы при условии полученных данных, которая вовсе недоступна p -значению! Приведем формулу:

$$P(H_k|D) = \frac{BF_k}{BF_k + 1} \quad (7)$$

Так как байесовский фактор может быть рассчитан двумя способами, в уравнении (7) мы использовали индексы k для обозначения гипотезы и байесовского фактора, увеличение которого свидетельствует в ее пользу. Например, для определения $P(H_0|D)$ в формулу необходимо подставить BF_{01} .

Применение и интерпретация байесовского фактора. Внимательный читатель уже наверняка заметил, что психология, наука, имеющая чрезвычайно сложный, многомерный, полидетерминированный, динамичный, многократно опосредованный объект изучения – психику, только выигрывает от применения байесовского фактора. Не привязываясь к

уровням значимости, показателям p -значения, исследователь получает в руки инструмент, который позволяет ему оценивать правдоподобность его гипотез. Причем инструмент значительно более точно и тонко устроенный, чем p -значение. Это означает в том числе, что применение байесовского фактора не требует тех огромных выборок, которые возникают иногда в психологических исследованиях.

В то же время недостатком байесовского фактора является сложность его расчета. Вероятно, именно этим объясняется тот факт, что он стал находить широкое использование только в компьютерную эру. Собственно представление алгоритмов расчета не входит в цели этой статьи, мы отсылаем вас к работе [10]. Также удачно реализован алгоритм определения BF в статистическом пакете JASP, к достоинствам которого можно отнести еще и бесплатность распространения, и вывод всех результатов согласно требованиям APA (Американской психологической ассоциации).

Как же интерпретировать байесовский фактор? Согласно соотносительной природе этого показателя, его основная интерпретация очень проста – во сколько раз полученные данные более вероятны при условии истинности одной гипотезы, чем при условии истинности другой (в случае BF_{01} – во сколько раз данные более вероятны при условии истинности нулевой гипотезы по сравнению с альтернативной, BF_{10} – при альтернативной по сравнению с нулевой). Например, $BF_{10}=12$ указывает, что полученные данные в 12 раз более вероятны при условии истинности альтернативной гипотезы, чем при условии истинности нулевой. Преимущество байесовского фактора в исследовательской практике прекрасно иллюстрируют Э. Ярош и Дж. Уайли: «...вместо того, чтобы просто сказать: “Маловероятно, что нет никакой связи между этими переменными, – исследователь может произнести, – альтернативная модель значительно лучше, чем нулевая и у меня есть вероятности, чтобы доказать это”» [10, с. 7].

С другой стороны, помимо прямой математической, несколько громоздкой интерпретации байесовского фактора, разрабатывается интерпретация, основанная на концепте «доказательство» (С. Гудман, Х. Джеффрис, Р. Кас, А. Рафтери). Тогда в случае $BF_{10}=12$, мы можем сказать, что альтернативная гипотеза имеет в 12 раз больше доказательств в свою пользу, чем нулевая. В «доказательном» подходе также разработаны определенные качественные характеристики BF . Представим их в таблице 1 (таблица приводится на основе работы [10]).

Таблица 1

Качественная интерпретация байесовского фактора

| | | Доказательства в пользу H_1 | |
|-------------|-----------|-------------------------------|---------------------|
| BF_{01} | BF_{10} | По А. Рафтери | По Х. Джеффрису |
| 1 – 0,33 | 1–3 | Слабые | В отдельных случаях |
| 0,33–0,1 | 3–10 | Положительные | Существенные |
| 0,1–0,05 | 10–20 | Положительные | Сильные |
| 0,05–0,03 | 20–30 | Сильные | Сильные |
| 0,03–0,01 | 30–100 | Сильные | Очень сильные |
| 0,01–0,0067 | 100–150 | Сильные | Решающие |
| <0,0067 | >150 | Очень сильные | Решающие |

Тогда, получив $BF_{10}=12$, исследователь мог бы заключить: «Альтернативная гипотеза имеет положительные (по Рафтери) или сильные (по Джеффрису) доказательства в пользу своей истинности».

Также разработаны системы интерпретации логарифма байесовского фактора, который называется *весом доказательств*, и собственно апостериорной вероятности гипотезы. Приведем их в одной таблице: данные для веса доказательства приводятся по Р. Касу, А. Рафтери [12], для апостериорной вероятности гипотезы по А. Рафтери [14].

Таблица 2

Качественная интерпретация веса доказательств и апостериорной вероятности гипотезы

| $\log(\text{BF}_{10})$ | Доказательства в пользу H_1 | $P(H_1 D)$ | Доказательства в пользу H_1 |
|------------------------|------------------------------------|------------|-------------------------------|
| 0–0,5 | Не достойно больше, чем упоминания | 50–75 | Слабые |
| 0,5–1 | Существенные | 75–95 | Положительные |
| 1–2 | Сильные | 95–99 | Сильные |
| >2 | Решающие | >0,99 | Очень сильные |

Приведем пример нашего психологического исследования влияния фактора пола на особенности саморазвития личности, в котором использование p -значения могло привести к неверным выводам [1]. Обратимся к анализу различий между мужчинами и женщинами только по одной переменной – саморегуляция в структуре саморазвития. Для сравнения средних значений был использован t -критерий Стьюдента. Выборку составили 296 человек, в возрасте от 17 до 79 лет, среди них 82 мужчины и 214 женщин.

Были получены такие результаты: среднее значение переменной саморегуляция у мужчин $\overline{CP}_m = -0,17$, у женщин – $\overline{CP}_ж = 0,07$. Данные всей выборки были представлены в стандартном нормальном распределении ($\bar{X}=0$, $s=1$). Итак, просто описательно мы видим, что женщины несколько превосходят мужчин по уровню саморегуляции. Значение критерия Стьюдента $t = -1,8$, при $p=0,068$. Какой вывод должен сделать исследователь, руководствуясь фреквентистской логикой? Полученное $p_{эмп} > 0,05$, соответственно статистически значимых различий по уровню саморегуляции между мужчинами и женщинами нет или, говоря языком статистики, – верна нулевая гипотеза. «Постойте, но ведь $p=0,068$ очень близко к $p=0,05$, – так может думать его не столь верный фреквентистскому подходу коллега, – значит, различия все-таки есть и верна скорее альтернативная гипотеза, женщины превосходят мужчин по уровню саморегуляции!». И каждый из

них будет строить свои объяснения, приводить доводы и доказательства, обращаться к теориям. А вся проблема в критическом уровне $p < 0,05$.

Посмотрим на байесовские факторы: $BF_{10}=0,691$; $BF_{01}=1,448$; $\log(BF_{10})= -0,37$. Что получается? Мы имеем очень слабые доказательства в пользу альтернативной гипотезы о том, что между мужчинами и женщинами есть различия, но мы же имеем и только слабые доказательства в пользу нулевой гипотезы: ни BF_{10} , ни BF_{01} не достигли 3. Х. Джеффрис еще называет это «в отдельных случаях», то есть в отдельных случаях есть различия между мужчинами и женщинами по уровню саморегуляции, а в отдельных случаях нет их вовсе. Если проинтерпретировать логарифм байесовского фактора, понятно, что минус перед значением говорит «в пользу нулевой гипотезы», но на качественном уровне это все же значит «не достойны больше, чем упоминания», ведь $\log(BF_{10}) < |0,5|$.

Суть проблемы становится еще заметней, если перевести байесовские факторы в апостериорные вероятности гипотез по формуле (7). Получим $P(H_1/D)=0,41$ или 41%; $P(H_0/D)=0,59$ или 59%. Другими словами, вероятности гипотез близки к случайным 50/50, как если бы вы подбрасывали монетку много раз и фиксировали, когда она выпадает орлом, а когда решкой. Говоря языком научной статистики: мы не имеем существенных доказательств, основанных на полученных данных, ни в пользу нулевой, ни в пользу альтернативной гипотезы. Вот такой вывод.

Выводы:

1) Привычное психологам p -значение имеет ряд существенных недостатков: связано с гипотетическими данными, зависит от намерений исследователя, от размера выборки, но самое неприятное – не предоставляет статистических доказательств: ни в пользу истинности нулевой, ни, уж тем более, в пользу истинности альтернативной гипотез. p -значение – вероятность данных, полученных в исследовании при условии истинности нулевой гипотезы, но никак не вероятность истинности самой гипотезы.

2) Альтернатива p -значению – байесовский фактор или отношение правдоподобия. Не имея ни одного из названных недостатков p -значения, он представляет собой отношение вероятности данных при условии истинности одной гипотезы к вероятности данных при условии истинности другой. Значительно проще интерпретируется как оценка силы доказательств, полученных в исследовании, в пользу той или иной гипотезы.

Список использованных источников

1. Зімовін О. І. Особливості саморозвитку особистості у залежності від статевих та сімейних чинників / О. І. Зімовін // Матеріали Міжнародної науково-практичної конференції: «Зимові наукові читання», 3 частина. – К.: Центр наукових публікацій, 2016. – С. 70–74.
2. Морен Э. Метод. Природа Природы / Э. Морен. – М. : Прогресс-Традиция, 2005. – 464 с.
3. Хей Дж. Введение в методы байесовского статистического вывода / Дж. Хей. – М. : Финансы и статистика, 1987. – 335 с.
4. Cohen J. The Earth is Round ($p < .05$) / Jacob Cohen // American Psychologist. – 1994. – №12, Vol. 49. – P. 997–1003.
5. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions / S. Goodman // Seminars in Hematology. – 2008. – №45(3). – P. 135–140.
6. Goodman S. Basic Bayes: I [Электронный ресурс] / S. Goodman. – Режим доступа: <http://www.gemini-grp.com/Bayes/GoodmanTutorial.pdf>
7. Goodman S. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy / S. Goodman // Annals of Internal Medicine. – 1999. – №12, Vol. 130. – P. 995–1004.
8. Goodman S. Toward Evidence-Based Medical Statistics. 2: The Bayes Factor / S. Goodman // Annals of Internal Medicine. – 1999. – №12, Vol. 130. – P. 1005–1013.
9. Ioannidis J. Why most published research findings are false / John Ioannidis // PLoS Medicine. – 2005. – № 8, Vol. 2. – P. 696–701.

10. Jarosz A. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors / A. Jarosz, J. Wiley // *Journal of Problem Solving*. – 2014. – Vol. 7. – P. 2–9.
11. Kanazawa S. Violent Men Have More Sons: Further Evidence for the Generalized Trivers-Willard Hypothesis (gTWH) / Satoshi Kanazawa // *Journal of Theoretical Biology*. – 2006. – №239 (4). – P. 450–459.
12. Kass R. Bayes Factors / R. Kass, A. Raftery // *Journal of the American Statistical Association*. – 1995. – №430, Vol. 90. – P. 773–795.
13. Masson M. A Tutorial on a Practical Bayesian Alternative to Null-Hypothesis Significance Testing / Michael Masson // *Behavior Research Methods*. – 2011. – Vol. 43. – P. 679–690.
14. Raftery A. Bayesian Model Selection in Social Research / A. Raftery // *Sociological Methodology*. – 1995. – Vol. 25. – P. 111–163.
15. Trafimow D. Editorial, *Basic and Applied Social Psychology* 37:1, 1-2 [Электронный ресурс] / Trafimow David, Michael Marks. – Режим доступа: <http://dx.doi.org/10.1080/01973533.2015.1012991>
16. Wagenmakers E.-J. A Practical Solution to the Pervasive Problems of P Values / E.-J. Wagenmakers // *Psychonomic Bulletin and Review*. – 2007. – №14(5). – P. 779–804.

References transliterated

1. Zimovin O. I. Osoblivosti samorozvitku osobistosti u zalezhnosti vid statevikh ta simeynikh chinnikov / O. I. Zimovin // *Materiali Mizhnarodnoï naukovo-praktichnoï konferentsii: «Zimovi naukovy chitannya», 3 chastina*. – K.: Tsentr naukovikh publikatsiy, 2016. – S. 70–74.
2. Moren E. *Metod. Priroda Prirody* / E. Moren. – M. : Progress-Traditsiya, 2005. – 464 s.
3. Khey Dzh. *Vvedenie v metody bayesovskogo statisticheskogo vyvoda* / Dzh. Khey. – M. : Finansy i statistika, 1987. – 335 s.
4. Cohen J. The Earth is Round ($p < .05$) / Jacob Cohen // *American Psychologist*. – 1994. – №12, Vol. 49. – P. 997–1003.

5. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions / S. Goodman // *Seminars in Hematology*. – 2008. – №45(3). – P. 135–140.
6. Goodman S. Basic Bayes: I [Электронный ресурс] / S. Goodman. – Режим доступа: <http://www.gemini-grp.com/Bayes/GoodmanTutorial.pdf>
7. Goodman S. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy / S. Goodman // *Annals of Internal Medicine*. – 1999. – №12, Vol. 130. – P. 995–1004.
8. Goodman S. Toward Evidence-Based Medical Statistics. 2: The Bayes Factor / S. Goodman // *Annals of Internal Medicine*. – 1999. – №12, Vol. 130. – P. 1005–1013.
9. Ioannidis J. Why most published research findings are false / John Ioannidis // *PLoS Medicine*. – 2005. – № 8, Vol. 2. – P. 696–701.
10. Jarosz A. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors / A. Jarosz, J. Wiley // *Journal of Problem Solving*. – 2014. – Vol. 7. – P. 2–9.
11. Kanazawa S. Violent Men Have More Sons: Further Evidence for the Generalized Trivers-Willard Hypothesis (gTWH) / Satoshi Kanazawa // *Journal of Theoretical Biology*. – 2006. – №239 (4). – P. 450–459.
12. Kass R. Bayes Factors / R. Kass, A. Raftery // *Journal of the American Statistical Association*. – 1995. – №430, Vol. 90. – P. 773–795.
13. Masson M. A Tutorial on a Practical Bayesian Alternative to Null-Hypothesis Significance Testing / Michael Masson // *Behavior Research Methods*. – 2011. – Vol. 43. – P. 679–690.
14. Raftery A. Bayesian Model Selection in Social Research / A. Raftery // *Sociological Methodology*. – 1995. – Vol. 25. – P. 111–163.
15. Trafimow D. Editorial, *Basic and Applied Social Psychology* 37:1, 1-2 [Электронный ресурс] / D. Trafimow, M. Marks. – Режим доступа: <http://dx.doi.org/10.1080/01973533.2015.1012991>
16. Wagenmakers E.-J. A Practical Solution to the Pervasive Problems of P Values / E.-J. Wagenmakers // *Psychonomic Bulletin and Review*. – 2007. – №14(5). – P. 779–804.