УДК 519.23:303.714:004.43:616.211-002.193-056.3
<https://doi.org/10.18692/1810-4800-2020-5-14-24>

Статистический анализ и оценка клинического эффекта на основе фактора Байеса в оториноларингологии

А. А. Корнеенков¹, О. И. Коноплев¹, И. В. Фанта¹, С. В. Левин¹, Е. Э. Вяземская¹

¹ Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи,
Санкт-Петербург, 190013, Россия

Statistical analysis and estimation of clinical effect based on Bayes factor for otorhinolaryngology

A. A. Korneenkov¹, O. I. Konoplev¹, I. V. Fanta¹, S. V. Levin¹, E. E. Vyazemskaya¹

¹ Saint Petersburg Research Institute of Ear, Throat, Nose and Speech,
Saint Petersburg, 190013, Russia

В статье рассматривается использование байесовских методов для статистического вывода как альтернативного традиционному методу проверки гипотез на основе уровня значимости. Приведены наглядные примеры решения традиционных статистических задач проверки гипотез в оториноларингологии на основе расчета и интерпретации фактора Байеса. В качестве двух задач для иллюстрации были использованы задача оценки влияния посещения сауны на показатели назального потока у больных аллергическим ринитом и задача оценки ассоциации сезона года и частоты рождения детей с нарушениями слуха. Хотя в форме статьи невозможно полностью описать и объяснить все математические термины и их происхождение для понимания логики байесовских методов, мы попытались объяснить, что они означают без отсылки к математическим руководствам. Поскольку байесовские методы все чаще используются в статистических приложениях, базовое понимание того, как их вычислять, должно быть частью инструментария каждого медицинского исследователя, а как интерпретировать – каждого практикующего врача, интересующегося современными результатами клинических исследований. Все расчеты, использованные в статье, сопровождаются R-кодом, поэтому они легко могут быть воспроизведены, текст статьи может быть использован как пошаговая инструкция для их выполнения.

Ключевые слова: клинические исследования, размер эффекта, уровень значимости, фактор Байеса, проверка статистических гипотез, язык R, аудиологический скрининг, аллергический ринит.

Для цитирования: Корнеенков А. А., Коноплев О. И., Фанта И. В., Левин С. В., Вяземская Е. Э. Статистический анализ и оценка клинического эффекта на основе фактора Байеса в оториноларингологии. *Российская оториноларингология*. 2020;19(5):14–24. <https://doi.org/10.18692/1810-4800-2020-5-14-24>

The article discusses the use of Bayesian methods for statistical inference as an alternative to the traditional method of testing hypotheses based on the significance level. Illustrative examples of solving traditional statistical problems of hypothesis testing in otorhinolaryngology based on the calculation and interpretation of the Bayes factor are presented. As two tasks for illustration, we used the tasks of assessing the impact of sauna visits on indicators of nasal flow in patients with allergic rhinitis and the task of assessing the association of the season of the year and the frequency of birth of children with hearing impairment. Although in the form of an article it is impossible to fully describe and explain all the mathematical terms and their origin for understanding the logic of Bayesian methods, we tried to explain what they mean without reference to the mathematical manuals. As Bayesian methods are increasingly used in statistical applications, a basic understanding of how to calculate them should be part of the toolkit of every medical researcher, and how to interpret it, of every practitioner who is interested in modern results of clinical trials. All calculations used in the article are accompanied by an R-code, so they can easily be reproduced, the text of the article can be used as step-by-step instructions for their implementation.

Keywords: clinical trials, effect size, significance level, Bayes factor, statistical hypothesis testing, R language, audiological screening, allergic rhinitis.

For citation: Korneenkov A. A., Konoplev O. I., Fanta I. V., Levin S. V., Vyazemskaya E. E. Statistical analysis and estimation of clinical effect based on Bayes factor for otorhinolaryngology. *Rossiiskaya otorinolaringologiya*. 2020;19(5):14–24. <https://doi.org/10.18692/1810-4800-2020-5-14-24>

Введение

В ходе клинических исследований возникает задача доказательства предположений, которые в обычном языке обозначаются словом «гипотеза». Проверка гипотезы может быть осуществлена двумя способами: прямым (логически или путем эксперимента) и косвенным. Провести эксперимент на живом человеке, поместить какой-либо орган, ткань, систему и т. п. в специально созданные условия редко представляется возможным. Поэтому если прямой способ отсутствует или невозможен, то прибегают к косвенной проверке гипотез. Косвенная проверка состоит в анализе следствий, которые логически вытекают из содержания гипотезы. Например, гипотеза может быть признана неверной, если некоторое явление, которое логически неизбежно следует из гипотезы, в клинической практике не наблюдается или, если происходит то, что при справедливости гипотезы происходить не должно. Результат косвенной проверки гипотез всегда дихотомичен: гипотеза признается либо верной, либо неверной.

Если гипотеза проверяется на основе собранных статистическим путем выборочных данных, то имеет место статистическая проверка гипотезы. Она состоит в выяснении вероятности того, насколько совместима выдвинутая гипотеза с наблюдаемым результатом. Конструкция выдвигаемой (или основной) статистической гипотезы, следуя логике косвенной проверки, должна быть противоположна тому факту или утверждению, которые необходимо доказать. Сформулированная по указанному принципу гипотеза называется нулевой и обозначается как H_0 ; противоположная гипотеза называется альтернативной и обозначается H_1 .

Все статистические тесты проверки гипотез (статистическая проверка гипотезы) проводятся на основе статистических моделей – статистических предположений (или набора статистических предположений), которые позволяют вычислять вероятность любого события и сгенерировать данные для выборки или всей популяции. По сути, сформулированные нулевая и альтернативная гипотезы представляют собой разные статистические модели, которые могут генерировать наборы данных, соответствующие этим гипотезам. Нулевая гипотеза является позицией по умолчанию, что не существует никакой связи между двумя наборами данных или нет ассоци-

ации между группами, например подвергшихся лечению или нет, а если и наблюдаются какие-либо отклонения, то они связаны со случайной изменчивостью признака. Нулевая гипотеза, как правило, считается верной до тех пор, пока доказательства не укажут обратное (аналогично «презумпции невиновности», когда обвиняемый считается невиновным до тех пор, пока его вина не будет доказана). Уровень значимости статистического теста (p -level) – допустимая для данной задачи вероятность ошибки первого рода (ложноположительного решения, false positive), то есть вероятность отклонить нулевую гипотезу, когда на самом деле она верна. Другими словами, уровень значимости дает возможность считать наблюдаемую связь между наборами данных либо случайной, либо нет.

Многие исследователи [1] указывают на проблемы неконкретности гипотезы H_1 . Если статистические предположения в H_0 конкретны: разница в параметрах между исследуемыми группами «равна 0», то гипотеза H_1 , как правило, представляет собой просто отрицание нулевой гипотезы: разница «не равна 0». Но «не равно 0» может быть и 1 и 1000 и т. д. Альтернативная основа для проверки статистических гипотез состоит в том, чтобы указать набор статистических моделей, по одной для каждой гипотезы, а затем использовать специальные методы (критерий Акаике, фактор Байеса и др.) для выбора наиболее подходящей модели.

Фактор Байеса (англ. Bayes factor) представляет собой количественное выражение преимущества одной модели по сравнению с другой моделью, независимо от того, верны модели или нет [2]. В последние годы использование статистических методов на основе фактора Байеса становится все более широким, появился даже термин «байесовцы» («Bayesians»), описывающий исследователей, предпочитающих байесовские методы традиционным способам проверки гипотез. Учитывая, что современные руководства, описывающие статистические принципы проведения клинических испытаний (например, CPMP/ICH/363/96 [3]), указывают, что использование байесовских и других подходов может иметь место, если обоснована их целесообразность, логично было бы использовать, а не игнорировать те преимущества, которые эти подходы имеют при решении задач клинических исследований.

Цель исследования

Изучение и описание статистических методов на основе фактора Байеса для решения традиционных задач оториноларингологии и иллюстрация их пошагового применения на языке R. Все R-коды для генерации и обработки данных содержатся в подразделе «Файлы для скачивания» раздела «Публикации» в файле bf.R (<https://lornii.ru/press-centr/publikatsii/bf.R>).

Материалы и методы исследования

Медицинским специалистам, особенно диагностам, очень легко понять смысл фактора Байеса. Подобный коэффициент обязателен к использованию (согласно приложению «А» ГОСТ Р 52600.0–2006 «Протоколы ведения больных. Общие положения») при описании диагностических медицинских технологий. Это отношение правдоподобия (Likelihood ratio) положительного (LR+) или отрицательного (LR–) результатов диагностического теста. Хотя есть некоторые нюансы в интерпретации, суть этих коэффициентов похожа – это отношение вероятности определенного клинического исхода или результата теста при двух разных статистических предположениях, гипотезах [4].

Для иллюстрации логики байесовского анализа на конкретном примере предположим, что проводится исследование клинической эффективности местной анестезии в носовых тампонах на определенный час после септопластики. Сформирована исследуемая группа, в которой применялась местная анестезия, и контрольная группа с плацебо (раствор хлорида натрия). Оценка болевых ощущений проводилась по вербальной аналоговой шкале (VAS) традиционно от 0 и 10 баллов. Размер эффекта (Effect Size, ES) оценивался в единицах стандартизованной средней разности (*d*-Козна, SMD): средний – 0,5; большой – 0,8. Методика оценки размера эффекта подробно описана в нашей работе [5]. Предполагалось, что часть авторов, соответствующих консервативному большинству исследователей этого вопроса (произвольно назовем ее группой С, от control), склонялась к тому, чтобы оценивать размер эффекта как средний (SMD = 0,5), часть авторов, считающих, что эффект более выражен (группа Т, от treatment) – как большой (SMD = 0,8). Представим, что нами было проведено дополнительное клиническое исследование, в котором размер эффекта оказался равным 0,48. Гипотезы двух групп (С и Т) авторов можно рассматривать как априорные (т. е. до нашего клинического исследования) суждения, или назовем их «приорами» (в англоязычной литературе они обозначаются коротко – prior). Наш результат ближе к гипотезе, выдвигаемой в группе С, однако он не противоречит и гипотезе, выдвиг-

нутой учеными группы Т. Вопрос заключается в том, какой гипотезе добавляет доказательности наше исследование? Начнем с идеи *априорных* (prior) шансов, описывающих степень, в которой мы предпочитаем одну гипотезу перед другой, их соотношение, прежде чем мы проведем собственное исследование:

$$\frac{P(H_t)}{P(H_c)},$$

где P – вероятность (или правдоподобие), а H_c и H_t являются гипотезами группы С и группы Т. Мы также можем описать *апостериорные* (posterior) шансы, описывающие степень, в которой мы предпочитаем одну гипотезу перед другой после проведения нашего исследования:

$$\frac{P(H_t|y)}{P(H_c|y)},$$

где y – это тот размер эффекта, что мы наблюдали. Теперь вопрос заключается в том, как правильно перейти от априорных шансов к апостериорным.

Правило Байеса гласит, что отношение правдоподобия двух гипотез изменяется в результате нашего исследования и получения новых данных определенным образом:

$$\frac{P(H_t|y)}{P(H_c|y)} = \frac{P(y|H_t)}{P(y|H_c)} \times \frac{P(H_t)}{P(H_c)}.$$

Коэффициент в середине:

$$\frac{P(y|H_t)}{P(y|H_c)}$$

есть фактор, на который мы умножаем априорные шансы, чтобы получить апостериорные шансы. Он показывает, во сколько раз изменится соотношение вероятностей двух гипотез после получения наших данных, и называется фактором Байеса.

Наши расчеты показали (R-код приведен в файле bf.R), что наблюдать размер эффекта $ES = 0,48$ более вероятно (с вероятностью 0,063) при истинности гипотезы Т, чем при истинности гипотезы С (0,004) (рис. 1.). По оси ординат указано значение функции плотности распределения вероятности случайной величины (Density, или ФПРВ), по оси абсцисс – значение случайной величины – наблюдаемого размера эффекта (Observed Effect Size). Мы не будем акцентировать внимание на то, почему именно так выглядят кривые ФПРВ для ES двух гипотез, предполагаем, что все параметры для рисования кривых распределений ES были получены группами исследователей С и Т. Обычно фактор Байеса обозначается BF (Bayes factor), а индексы в виде подстрочных символов обозначают сравниваемые гипотезы, например, BF_{10} – это фактор Байеса BF, а индекс 1 означает гипотезу-числитель H_1 и 0 – гипотезу-знамена-

Рис. 1. Вертикальная линия соответствует наблюдаемому значению ES, высота P_t соответствует вероятности (или значению ФПРВ) этого значения по гипотезе H_t , а P_c – по гипотезе H_c . Фактор Байеса – это отношение этих вероятностей для наблюдаемого значения ES ($BF_{tc} \approx 16$)

Fig. 1. The vertical line corresponds to the observed value ES, the height P_t corresponds to the probability (or the value of the density) of this value under the hypothesis H_t , and P_c corresponds to the hypothesis H_c . The Bayes factor is the ratio of these probabilities for the observed value ES ($BF_{tc} \approx 16$)

тель H_0 в формуле расчета фактора Байеса. Для нашего примера по этой логике фактор Байеса обозначен как BF_{tc} .

Фактор Байеса может быть любым положительным числом. Кроме того, фактор Байеса может быть интерпретирован в номинальной шкале, определенным прилагательным, указывающим на силу доказательств («сильный», «слабый» и т. п.). Одна из наиболее распространенных интерпретаций, впервые предложенная Гарольдом Джеффрейсом [6] и слегка измененная С. Андрасевич [1], представлена в табл. 1.

На рис. 2 показана зависимость величины фактора Байеса от наблюдаемого размера эффекта. Наблюдаемые данные благоприятствуют H_t , когда $BF > 1$, что происходит, когда ES больше 0,4.

Предпочтение следует отдать H_c , когда $BF < 1$, что происходит, когда H_c меньше 0,4. Максимальное значение BF при $ES = 0,76$, а затем по мере увеличения ES размер BF уменьшается.

Фактор Байеса интересен для проверки того, насколько наши данные делают предпочтительнее одну из двух конкретных гипотез, но этот инструмент может использоваться и при традиционной проверке статистических гипотез, когда H_1 – это просто отрицание H_0 без конкретного указания значения.

При использовании и интерпретации фактора Байеса есть важные особенности. Байесовский фактор можно рассматривать как меру доказательства в пользу нулевой гипотезы H_0 . Если $BF < 0,05$, то апостериорные шансы в пользу H_0 бу-

Таблица 1

Уровень (категории) доказательности для разных значений фактора Байеса

Table 1

Evidence category (level) for different Bayes factor values

Значение BF_{10} ...	Интерпретация
> 100	Экстремальные доказательства для H_1
30–100	Очень убедительные доказательства H_1
10–30	Сильные доказательства для H_1
3–10	Умеренные доказательства для H_1
1–3	Анекдотические доказательства для H_1
1	Нет доказательств
1/3–1	Анекдотические доказательства H_0
1/3–1/10	Умеренные доказательства H_0
1/10–1/30	Сильные доказательства H_0
1/30–1/100	Очень убедительные доказательства H_0
$< 1/100$	Экстремальные доказательства для H_0

Рис. 2. Зависимость величины фактора Байеса от наблюдаемого размера эффекта ($x = ES$).
Fig. 2. Bayes factor value dependent on observed effect size ($x = ES$).

дут меньше, чем одна двадцатая часть априорных шансов. Это является значительным доказательством против H_0 . Например, если у вас до вашего исследования разница между «за» и «против» H_0 была небольшая (шансы один к одному), а с учетом фактора Байеса, равного 0,05, апостериорные шансы также будут 0,05 (5 к 100 или 1 к 20). Мы можем легко конвертировать шансы обратно в вероятность, используя формулу: Вероятность = Шансы / (1 + Шансы). Апостериорные шансы менее 0,05 преобразуются в апостериорную вероятность $0,05/1,05 = 0,0476$. Таким образом, при значении фактора Байеса меньше 0,05 существует лишь небольшая (0,0476) апостериорная вероятность того, что H_0 истинна, если только не было априорных доказательств в ее пользу.

Фактор Байеса должен интерпретироваться относительно наших априорных представлений об исследуемом предмете. Например, фактор Байеса, равный 0,01, действительно является убедительным доказательством против H_0 , но если априорные доказательства были в ее пользу, например с априорными шансами 1000 (1000 к 1), то апостериорные шансы по-прежнему будут в пользу H_0 . Например, если бы H_0 была гипотезой, что человек не умеет читать мысли на расстоянии, то большинству людей потребуются чрезвычайно убедительные экспериментальные доказательства, чтобы преодолеть твердое априорное убеждение, что у людей нет такой экстрасенсорной способности.

Для иллюстрации байесовского теста на ассоциацию номинальных переменных и оценку влияния фактора на непрерывную переменную

мы используем программную среду R с несколькими программными пакетами (BayesFactor и BayesianFirstAid). Хотя эти задачи можно решить и встроенными средствами R, использование библиотек значительно упрощает получение результата и его визуализацию.

Задача 1. В исследовании [7] измерялся эффект у больных аллергическим ринитом от посещения в течение 6 недель сауны. Как известно, аллергический ринит является хроническим респираторным заболеванием, при котором у пациентов выявляется симпатическая гипопункция. Было изучено влияние планового посещения сауны на вегетативную нервную систему, PNIF (peak nasal inspiratory flow – пиковый назальный инспираторный поток) и функции легких с аллергическим ринитом. Диагноз аллергический ринит определялся на основании анамнеза, физического осмотра и скарификационной пробы. Участвующие в исследовании были случайным образом распределены на две группы. Участники группы контроля поддерживали обычную жизнь. Участники в исследуемой группе проходили лечение в сауне в течение шести недель – 3 дня в неделю, с 6 сеансами по 5 минут в день – всего 30 минут. В начале исследования различий по всем исследуемым показателям между группами обнаружено не было. Спустя 6 недель испытаний у пациентов были измерены различные показатели, в том числе PNIF. Среднее значение PNIF и стандартное отклонение (Standard Deviation, SD) в контрольной группе и в исследуемой группе составило 103,0 (31,3) L/min и 161,9 (46,7) L/min соответственно [8].

Таблица 2

R-код для генерации данных исследования

Table 2

R code for generating study data

№	R код
1.	PTrmu<-161.9 # средняя PNIF в группе лечения
2.	PCTmu<-103.0 # средняя PNIF в группе контроля
3.	PTrSD<-46.7 # SD PNIF в группе лечения
4.	PCTSD<-31.3 # SD PNIF в группе контроля
5.	n<-13 # число испытуемых в каждой группе
6.	set.seed(1234)
7.	df <- data.frame(GROUP=factor(rep(c("SAUN", "CTRL"), each=n)), pnif=round(c(rnorm(n, mean=PTrmu, sd=PTrSD), rnorm(n, mean=PCTmu, sd=PCTSD))))

На основе этих данных в среде R (R-код приведен в табл. 2) был сгенерирован набор данных аналогичным объемом, указанным в статье, с двумя переменными: 1) переменная группы пациентов с названием «GROUP», в которой значение «SAUN» обозначало пациентов исследуемой группы, в «CTRL» – пациентов контрольной группы; 2) переменная «pnif» с данными PNIF в L/min.

Научная гипотеза исследования заключалась в том, что посещение сауны по определенному плану может оказать эффект на показатель PNIF. Статистические гипотезы были сформулированы следующим образом:

$$H_0 : \text{pnif}_s = \text{pnif}_c$$

$$H_1 : \text{pnif}_s \neq \text{pnif}_c$$

где pnif_s и pnif_c – это значение PNIF в группе пациентов, посещающих сауну, и в группе, ее не посещавших, соответственно. Для решения подоб-

ных задач традиционно используется *t*-критерий Стьюдента или его непараметрическая альтернатива – критерий Манна–Уитни. Чтобы продемонстрировать различия в решении задачи и интерпретации результатов, она была решена двумя способами. Ниже приведена машинограмма (рис. 3), в которой сначала выполняется традиционный *t*-тест (мы сгенерировали данные нормально распределенными) с помощью функции *t.test()*, а затем расчет фактора Байеса с помощью функции *ttestBF()* из пакета *BaeyesFactor*.

Интерпретация результатов *t*-теста по машинограмме проста. Мы предполагаем, что дисперсии показателя в двух группах не различаются (*var.eq=TRUE*, что соответствует логике статистической проверки гипотез, по которой мы считаем наблюдаемые отклонения случайными до тех пор, пока не доказали обратное). В первой строке указано название теста «Two Sample *t*-test» (двухвыборочный *t*-тест), затем анализируемые

```
t.test(pnif~GROUP, data = df, var.eq=TRUE)
```

```
Two Sample t-test
```

```
data: pnif by GROUP
t = -2.8275, df = 24, p-value = 0.009314
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -70.12905 -10.94787
sample estimates:
mean in group CTRL mean in group SAUN
 99.61538          140.15385
```

```
bf = ttestBF(formula = pnif~GROUP, data = df)
```

```
bf
Bayes factor analysis
```

```
-----
[1] Alt., r=0.707 : 5.410284 ±0%
```

```
Against denominator:
  Null, mu1-mu2 = 0
---
```

```
Bayes factor type: BFindepSample, JZS
```

Рис. 3. Машинограмма выполнения традиционного *t*-теста с помощью функции *t.test()*, а затем расчет фактора Байеса с помощью функции *ttestBF()*

Fig. 3. The program code for performing the traditional *t*-test with the function *t.test()* and then Bayes factor with the *ttestBF()* function

данные (pnif by GROUP), полученные значения t -критерия Стьюдента (t), число степеней свободы (df) и уровень значимости p (p -value). Далее указываются формулировка альтернативной гипотезы H_1 (alternative hypothesis: true difference in mean is not equal to 0), 95%-ный доверительный интервал для истинного значения разницы средних (–70,12905; –10,94787), выборочные оценки средних в сравниваемых первой (CTRL) и второй (SAUN) группах, 99,61538 и 140,15385 соответственно. Так как уровень значимости p -value = 0,009314 и он менее обычно устанавливаемой ошибки первого рода $\alpha = 0,05$, можно считать, что произошло статистически значимое изменение PNIF после 6 недель сауны по сравнению с контрольной группой.

Представленный далее в машинограмме результат расчета фактора Байеса менее привычен и требует пояснения. Часть BFindepSample указывает, что проведен t -тест для независимых выборок, а «JZS» указывает на одну из методик вычисления (t -тест Jeffreys–Zellner–Siow (JZS) [9]. В строке выше, текст «Against denominator: Null, mul-mu2 = 0» указывает, что нулевая гипотеза, выступающая знаменателем в расчете фактора Байеса, состоит в том, что не существует различий между средними (μ). Строка выше указывает на числитель – альтернативную гипотезу (Alt.): [1] Alt., $r = 0,707$: 5,410284 $\pm$ 0%.

Указанный коэффициент $r = 0,707$ многими исследователями относится к технической части вывода и игнорируется. Однако и он имеет свой смысл. Известно, что фактор Байеса по умолчанию для t -тестов [9] предполагает, что эффект (Effect size (ES), выраженный через d Коэна (мы подробно рассматривали меры эффекта клинических воздействий в статье А. А. Корнеевкова, И. В. Фанты [5]), имеет распределение Коши (Cauchy), это распределение приора для H_1 . Распределение Коши описывает отношение двух нормально распределенных случайных величин и является частным случаем распределения Стьюдента. Внешне оно напоминает нормальное распределение, с его колоколообразной формой кривой функции плотности распределения вероятности, но имеет более «тяжелые», опущенные вниз «хвосты» этой кривой. Для описания распределения используются коэффициенты сдвига (обычно в функциях обозначается location) и масштаба (scale). Разброс, «размытие» распределения Коши можно изменить с помощью параметра масштаба r . В зависимости от конкретной области исследования можно использовать более «широкое» (большие r , например $r = 1,5$) или более «тонкое» (маленькие r , например $r = 0,5$) распределение Коши. Это соответствует предварительному убеждению, какой эффект следует ожидать, больший или меньший.

Для t -теста с двумя выборками пакет BayesFactor для R предлагает три значения по умолчанию для параметра масштаба: medium, средний ($r = 0,71$) (значение по умолчанию), wide, широкий ($r = 1$) и ultra-wide, ультраширокий ($r = 1,41$). Масштаб контролирует, насколько велики в среднем ожидаемые истинные размеры эффекта. В каждом из этих масштабов 50% истинных размеров эффекта находятся в интервале ($-r, r$). Для шкалы по умолчанию «средний» 50% мы предполагаем, что размеры эффекта находятся в пределах диапазона ($-0,7071; 0,7071$), для широкого – ($-1; 1$) и т. д. Увеличение r увеличивает размеры ожидаемых эффектов; уменьшение r уменьшает размер ожидаемых эффектов.

После указания заданного значения масштаба r следует фактор Байеса: доказательства, представленные наблюдаемыми данными, дают 5,410284: 1 (или округленно 5 : 1) в пользу альтернативной гипотезы, т. е. они указывают, что альтернативная гипотеза предпочтительнее нулевой гипотезы, что может быть интерпретировано по табл. 1 как «умеренные доказательства для H_1 ».

Задача 2. В РФ аудиологический скрининг новорожденных осуществляется в соответствии с Приказом Министерства здравоохранения и социального развития Российской Федерации от 28.04.2007 № 307 «О стандарте диспансерного (профилактического) наблюдения ребенка в течение первого года жизни», п. 6 «Диагностика и оценка функционального состояния организма». Согласно информации о количестве детей, обследованных на нарушение слуха в рамках проведения универсального аудиологического скрининга (АС) в 2019 году в одной из областей Северо-Западного федерального округа России, число новорожденных, выявленных с нарушениями слуха из обследованных в родильном доме (1а этап) и в детской поликлинике (1б этап) составило 267 человек из 7352 человек, обследованных на 1а и 1б этапах АС.

Было обращено внимание на то, что относительная частота нарушений слуха, выявленных у новорожденных в летние месяцы, несколько ниже, чем в другие времена года. Ниже приведена табл. 3 (таблица сопряженности), в которой указано число новорожденных с нарушением слуха и без него в летний период и другое время года.

Предположение об ассоциации частоты нарушения слуха у новорожденных и временем года может быть проверено статистически на основе сформулированных гипотез:

$$H_0: \theta_1 = \theta_2;$$

$$H_1: \theta_1 \neq \theta_2,$$

где θ_1 – относительная частота нарушений слуха в первой группе (летние месяцы); θ_2 – отно-

Таблица 3

Число новорожденных с нарушением слуха и без него в летний период и другое время года

Table 3

Number of hearing-impaired newborns in summer and at other times of the year

Время года	Нарушение слуха	
	Есть	Нет
Лето (Group 1)	53	1736
Другое время года (Group 2)	214	5349

сительная частота нарушений слуха во второй группе (остальные времена года). Group 1 – новорожденные в летние месяцы года. Group 2 – новорожденные в остальные месяцы года.

Для решения этой задачи был использован пакет BayesianFirstAid (Rasmus Bååth, см. <http://sumsar.net/blog/2014/01/bayesian-first-aid/>). Так же как и в задаче 1, статистический вывод был получен сначала с использованием традиционной статистики (для бинарных переменных это – Хи-квадрат), а затем с помощью ее байесовской альтернативы (рис. 4).

В начале R-кода определяется исходный набор данных для анализа, затем указывается название теста – двухвыборочный тест на равенство про-

порций с коррекцией на непрерывность. Далее приводится значение Хи-квадрат, числа степеней свободы (df) и уровень значимости p -level, значение которого (0,09561) больше критического допустимого значения ($\alpha = 0,05$) вероятности ошибки принять неверную альтернативную гипотезу. Поэтому верной признается нулевая гипотеза, которая говорит об отсутствии статистически значимого различия пропорций (относительных частот) нарушений слуха у родившихся в летние месяцы и в остальное время года. Далее указаны 95%-ный доверительный интервал для разности относительных частот в двух группах –0,01 (–0,017; 0,0013) и выборочные оценки пропорций в группе 1 (0,02962549) и группе 2 (0,03846845).

```
> deafs <- matrix(c(53,214,1736,5349),
+                 nrow = 2,
+                 dimnames = list(c("Summer", "Not Summer"), #row
+                               c("Deaf", "Not Deaf"))) # column
> # вывод матрицы
> deafs
      Deaf Not Deaf
Summer    53   1736
Not Summer 214   5349

> prop.test(deafs) # аналогичный результат при выполнении встроенной функции chisq.test(deafs)

      2-sample test for equality of
      proportions with continuity
      correction

data:  deafs
X-squared = 2.7772, df = 1, p-value =
0.09561
alternative hypothesis: two.sided
95 percent confidence interval:
-0.01855424  0.00086831
sample estimates:
prop 1      prop 2
0.02962549 0.03846845

> bayes.prop.test(deafs)

      Bayesian First Aid proportion test

data:  deafs
number of successes:    53, 214
number of trials:      1789, 5563
Estimated relative frequency of success [95% credible interval]:
  Group 1: 0.030 [0.022, 0.038]
  Group 2: 0.039 [0.034, 0.044]
Estimated group difference (Group 1 - Group 2):
-0.01 [-0.017, 0.0013]
The relative frequency of success is larger for Group 1 by a probability
of 0.042 and larger for Group 2 by a probability of 0.958 .
```

Рис. 4. Машинограмма выполнения традиционного χ^2 -теста с помощью функции `prop.test()`, а затем расчет фактора Байеса с помощью функции `bayes.prop.test()`

Fig. 4. Listing of the performing the traditional chi-square test with the `prop.test()` function, and then calculating the Bayes factor with the `bayes.prop.test()` function

В блоке результатов байесовского теста пропорций первые три строки: «data», «number of successes», «number of trials» повторяют данные, которые указаны как исходные. Четвертая строка представляет оценки относительных частот нарушений слуха (это интересующее нас событие обозначается в терминах статистических испытаний успехом – success) для двух групп: точечные (для группы 1: 0,030, для группы 2: 0,039) и интервальные (95%-ный доверительный интервал) [0,023, 0,038] и [0,034, 0,044] соответственно. Ниже указана оцениваемая разница этих частот для двух групп –0,01 [–0,017, 0,0013].

Последняя строка самая интересная, она указывает, что относительная частота нарушений слуха больше в группе 1, чем в группе 2 с вероятностью 0,041 и, наоборот, относительная частота нарушений слуха больше в группе 2, чем в группе 1 с вероятностью 0,959.

Пакет BaeyesFactor также предлагает свое решение этой задачи с помощью функции contingencyTableBF() (рис. 5).

После названия «Bayes factor analysis» приводится формулировка альтернативной гипотезы [«non-indep.(endence)» – «не независимость», т. е. зависимость ассоциации между переменными]. Указано значение фактора Байеса (0,05772486), его точность в виде процентов.

Выражение « $a = 1$ » несет, так же как и параметр масштаба r из первой задачи, серьезную смысловую нагрузку. Это так называемый параметр концентрации, который индексирует ожидаемое отклонение H_1 от H_0 и соответствует параметру концентрации a распределения Дирихле (Dirichlet), называемым часто «распределением распределений», поскольку его можно рассматривать как распределение самих вероятностей. Установка $a = 1$ является выбором по умолчанию, поскольку она соответствует предположению, что у нас нет предварительной информации, чтобы отдать предпочтение одному компоненту в этом распределении перед любым другим. По этому вопросу можно найти достаточно хорошее объяснение этого параметра [https://stats.stackexchange.com/q/244946].

С точки зрения интерпретации, $BF = 0,05$ означает, что наблюдаемые данные в $1/0,05 = 20$ раз более вероятны для гипотезы H_0 , чем для

H_1 [10]. Как уже указывалось выше, когда мы предполагаем, что конкурирующие модели одинаково вероятны априори (то есть когда априорные шансы равны 1), байесовский фактор может быть преобразован в апостериорную вероятность путем деления байесовского фактора на байесовский фактор плюс 1; например, при равной априорной вероятности коэффициент Байеса $BF_{12} = 0,05$ приводит к апостериорной вероятности, равной $0,05/(0,05+1) \approx 0,0476$.

В заключении необходимо остановиться на моделировании апостериорных вероятностей после получения значений фактора Байеса. Мы уделяем этому некоторое внимание, потому что при решении задач байесовскими методами практически всегда на основе полученных данных генерируются апостериорная вероятность и диаграммы, которые не всегда понятны неискушенному в математике медицинскому исследователю. Полученные выше значения апостериорных вероятностей – это не сами вероятности, а всего лишь их оценки. Например, если в нашем исследовании, в нашей выборке из 10 пациентов наблюдались осложнения у 2 человек, то такой исход может наблюдаться и при вероятности осложнения в популяции 50%, и при 10%. Но при вероятности в 20% (или правильнее 0,2) он будет наиболее вероятен или правдоподобен. Чтобы получить точное значение вероятности осложнения, нужно провести несколько выборочных исследований, но можно также использовать компьютерные реализации специальных методов их генерации, например метод Монте Карло по схеме марковской цепи (Markov Chain Monte-Carlo, MCMC). MCMC является набирающим популярность методом получения информации о распределениях, особенно для оценки апостериорных распределений в байесовском выводе [11]. Название MCMC сочетает в себе два свойства: цепь Монте-Карло и цепь Маркова. Монте-Карло – это метод оценки свойств распределения путем изучения случайных выборок из распределения. Например, вместо того чтобы находить среднее значение нормального распределения путем непосредственного вычисления его по уравнениям распределения, подход Монте-Карло будет состоять в том, чтобы извлечь большое количество случайных выборок из нормального

```
> bf = contingencyTableBF(deafS, sampleType = "indepMulti", fixedMargin = "rows")
> bf
Bayes factor analysis
-----
[1] Non-indep. (a=1) : 0.05772486 ±0%

Against denominator:
  Null, independence, a = 1
---
Bayes factor type: BFcontingencyTable, independent multinomial
```

Рис. 5. Машинограмма с расчетом фактора Байеса в пакете BaeyesFactor
Fig. 5. Listing of the calculating the Bayes factor in Baeyesfactor package

Рис. 6. Графическое представление результатов сравнения относительных частот нарушения слуха у новорожденных, диагностированных в летний период времени и в другое время года (N – число сгенерированных выборок, итераций; Bandwidth – ширина диапазонов группировки результатов).

Fig. 6. The graph presents data showing Change in probability of deaf birth in summer compared to other seasons (N is the number of generated samples, iterations; Bandwidth is range width of the results grouping ranges)

Таблица 4

R-код для генерации диаграммы рис. 6

Table 4

R-code for generating diagram Fig. 6

№	R-код
1.	<code>chains = posterior(bf, iterations = 10000)</code>
2.	<code>SumDeafs = chains[,"pi[1,1]"] / chains[,"pi[1,*]"]</code>
3.	<code>NoSumDeafs = chains[,"pi[2,1]"] / chains[,"pi[2,*]"]</code>
4.	<code>plot(mcmc(SumDeafs - NoSumDeafs), main = "Change in probability of deaf birth in summer compared to other seasons")</code>

распределения и вычислить среднее значение выборки из них. Свойство цепочки Маркова в МСМС состоит в том, что случайные выборки генерируются специальным последовательным процессом. Каждая случайная выборка используется в качестве ступеньки для создания следующей случайной выборки. Хотя каждая новая выборка зависит от предыдущей, новые выборки не зависят от каких-либо выборок до предыдущего (это «свойство Маркова»).

На рис. 6 представлены диаграммы с симулированными с помощью R-кода (табл. 4) значениями апостериорной вероятности разности относительных частот глухоты у родившихся в разное время года с помощью разных пакетов R. Второе изображение иллюстрирует то, как распределена вероятность наблюдать определенные значения разности частот рождения глухих в разные сезоны года, сгруппированные по небольшим

диапазонам (диапазон, определяется Bandwidth). Это соответствует значениям на машинограмме рис. 4 – $-0,01$ ($-0,017; 0,0013$).

Строки R-кода в табл. 4 генерируют рис. 6.

Выводы

Байесовские факторы являются важной частью прикладного инструментария байесовской статистики. Хотя в форме статьи невозможно описать технические детали того, как они получены или вычислены на практике, мы попытались объяснить, что они означают. Поскольку байесовские методы все чаще используются в статистических приложениях, базовое понимание того, как их интерпретировать, должно быть частью инструментария каждого ученого.

Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА/REFERENCES

- Andraszewicz S., Scheibehenne B., Rieskamp J., Grasman R., Verhagen J., Wagenmakers E.-J. An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*. 2015;41(2):521–543. <https://doi.org/10.1177/0149206314560412>
- Ly A., Verhagen J., Wagenmakers E.-J. Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*. 2016;72:19–32. <https://doi.org/10.1016/j.jmp.2015.06.004>

3. Note for guidance on statistical principles for clinical trials (CPMP/ICH/363/96) / Step 5. ICH Topic E 9 Statistical Principles for Clinical Trials. European Medicines Agency: September 1998 CPMP/ICH/363/96. EMEA. 2006. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf
4. Корнеев А. А., Рязанцев С. В., Вяземская Е. Э., Будкова М. А. Меры информативности диагностических медицинских технологий в оториноларингологии: вычисление и интерпретация. *Российская оториноларингология*. 2020;19(1):46–55 [Korneev A. A., Ryzantsev S. V., Vyazemskaya E. E., Budkovaya M. A. The measures of informativeness of diagnostic medical technologies in otorhinolaryngology: calculation and interpretation. *Rossiiskaya otorinolaringologiya*. 2020;19(1):46–55 (in Russ.)]. <https://doi.org/10.18692/1810-4800-2020-1-46-55>
5. Корнеев А. А., Фанта И. В. Оценка размера эффекта клинического воздействия. *Российская оториноларингология*. 2020;19(2):42–50 [Korneev A. A., Fanta I. V. Estimation of the effect size of clinical intervention in otorhinolaryngology. *Rossiiskaya otorinolaringologiya*. 2020;19(2):42–50 (in Russ.)]. <https://doi.org/10.18692/1810-4800-2020-2-42-50>
6. Jeffreys H. Theory of probability (3rd ed.). New York: Oxford University Press. 1961. 472 p.
7. Kunbootsri N., Janyacharoen T., Arrayawichanon P., Chainansamit S., Kanpittaya J., Auvichayapat P., Sawanyawisuth K. The effect of six-weeks of sauna on treatment autonomic nervous system, peak nasal inspiratory flow and lung functions of allergic rhinitis Thai patients. *Asian Pac J Allergy Immunol*. 2013;31(2):142-7. PMID: 23859414. <https://doi.org/10.12932/AP0262.31.2.2013>
8. Papachristou A., Bourli E., Aivazi D., Futzila E., Papastavrou T., Konstandinidis T., Maratou E., Ilonidis G., Aivazis V. Normal peak nasal inspiratory flow rate values in Greek children and adolescents. *Hippokratia*. 2008;12(2):94–97. PMID: 18923655; PMCID: PMC2464304. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2464304/>
9. Rouder J. N., Speckman P. L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*. 2009;16(2):225–237. <https://doi.org/10.3758/PBR.16.2.225>
10. Jamil T., Ly A., Morey R. D., Love J., Marsman M., Wagenmakers E. J. Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior research methods*. 2017;49(2):638–652. <https://doi.org/10.3758/s13428-016-0739-8>
11. Van Ravenzwaaij D., Cassey P., Brown S. D. A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic bulletin & review*. 2018;25(1):143–154. <https://doi.org/10.3758/s13423-016-1015-8>

Информация об авторах

✉ **Корнеев Алексей Александрович** – доктор медицинских наук, профессор, заведующий лабораторией информатики и статистики, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (904) 554-07-40, e-mail: korneyenkov@gmail.com
ORCID: <https://orcid.org/0000-0001-5870-8042>

Коноплев Олег Иванович – доктор медицинских наук, профессор, заместитель директора по клинической работе, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (812) 316-25-05, e-mail: oikonoplev@mail.ru
<https://orcid.org/0000-0002-2811-5462>

Фанта Иван Васильевич – кандидат медицинских наук, заведующий организационно-методическим отделом, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (812) 316-54-29, e-mail: 3165429@mail.ru
ORCID: <https://orcid.org/0000-0002-1110-7087>

Левин Сергей Владимирович – кандидат медицинских наук, старший научный сотрудник, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (812) 495-36-71, e-mail: sergeyln@mail.ru
ORCID: <https://orcid.org/0000-0001-9770-7739>

Вяземская Елена Эмильевна – инженер лаборатории информатики и статистики, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (911) 996-08-89, e-mail: vyazemskaya.elena@gmail.com
ORCID: <https://orcid.org/0000-0002-4141-2226>

Левина Елена Алексеевна – кандидат медицинских наук, старший научный сотрудник, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); тел.: +7 (812) 495-36-71, e-mail: xramoval@gmail.com
ORCID: <https://orcid.org/0000-0003-0285-6526>

Information about the authors

✉ **Alexei A. Korneev** – MD, Professor, Head of the Laboratory of Informatics and Statistics Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (904) 554-07-40, e-mail: korneyenkov@gmail.com

ORCID: <https://orcid.org/0000-0001-5870-8042>

Oleg I. Konoplev – MD, Professor, Deputy Director for Clinical Work, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (812) 316-25-05, e-mail: oikonoplev@mail.ru
ORCID: <https://orcid.org/0000-0002-2811-5462>

Ivan V. Fanta – PhD (Medicine), Head of the organizational and methodological department, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (812) 316-54-29, e-mail: 3165429@mail.ru
ORCID: <https://orcid.org/0000-0002-1110-7087>

Elena E. Vyazemskaya – engineer of the Laboratory of Informatics and Statistics, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (911) 996-08-89, e-mail: vyazemskaya.elena@gmail.com
ORCID: <https://orcid.org/0000-0002-4141-2226>

Sergei V. Levin – PhD (Medicine), Researcher, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (812) 495-36-71, e-mail: sergeyln@mail.ru
ORCID: <https://orcid.org/0000-0001-9770-7739>

Elena A. Levina – PhD (Medicine), Researcher, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya str., St. Petersburg, 190013, Russia); phone: +7 (812) 495-36-71, e-mail: xramoval@gmail.com
ORCID: <https://orcid.org/0000-0003-0285-6526>