

И. Н. Мороз, С. В. Власова

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ ПРИ ПРОВЕДЕНИИ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

УО «Белорусский государственный медицинский университет»,
Минск, Республика Беларусь

В статье представлены теоретические основы проведения медико-биологических исследований, включая важнейшие вопросы их планирования, связанные с проверкой статистических гипотез. Описаны классификация статистических гипотез, статистических критериев, этапы проверки гипотез, основанных на оценке статических критериев. Дана характеристика ошибкам 1-го и 2-го типа, статистической мощности, статическим критериям, которые могут использоваться для проверки гипотез о законах распределения данных. Показано, что выбор статического критерия проверки статистической гипотезы зависит от цели и задач исследования, дизайна, типа переменной и распределения данных, количества выборочных совокупностей и числа наблюдений в них, мощности самого статистического критерия, позволяющего выявлять различия или отклонять нулевую гипотезу.

Цель данной статьи заключается в попытке продемонстрировать необходимость проверки статистических гипотез на различных этапах проведения медико-биологических исследований, применяемых в доказательной медицине. Показана важность определения оптимального числа наблюдений изучаемой группы на этапе планирования исследования с учетом статистической мощности и ошибок 1-го и 2-го типа, обеспечивающие правомочность результатов проверки гипотез. Надеемся, что данная статья поможет исследователям осуществлять научно обоснованное планирование исследования, корректно использовать статистические критерии для анализа и интерпретации полученных результатов, основанных на проверке статистической гипотезы.

Ключевые слова: статистическая гипотеза, статистическая значимость, уровень значимости, ошибки 1-го и 2-го типа, статистическая мощность, параметрические и непараметрические критерии проверки гипотезы.

I. N. Moroz, S. V. Vlasova

STATISTICAL HYPOTHESES IN BIOMEDICAL RESEARCH

Belarusian State Medical University, Minsk, Republic of Belarus

The article presents the theoretical foundations of conducting medical and biological research, including the most important issues of their planning related to testing statistical hypotheses. The classification of statistical hypotheses, statistical criteria and stages of testing hypotheses based on the assessment of statistical criteria are described. The characteristics of type 1 and type 2 errors, statistical power, and statistical criteria that can be used to test hypotheses about the data distribution laws are given. It is shown that the choice of a statistical criterion for testing a statistical hypothesis depends on the purpose and objectives of the study, design, type of variable and data distribution, the number of sample populations and the number of observations in them, the power of the statistical criterion itself, which allows identifying differences or rejecting the null hypothesis.

The aim of this article is to attempt to demonstrate the need for testing statistical hypotheses at various stages of biomedical research used in evidence-based medicine. The importance of determining the optimal number of observations of the study group at the stage of research planning is shown, taking into account statistical power and errors of the 1st and 2nd types, ensuring the validity of the results of hypothesis testing. We hope that this article will help

researchers to plan the study, correctly use statistical criteria for the analysis and interpretation of the obtained results based on the testing of the statistical hypothesis.

Key words: *statistical hypothesis, statistical significance, confidence level, type 1 and type 2 errors, statistical power, parametric and nonparametric hypothesis testing criteria.*

Планирование медико-биологических исследований является неотъемлемой частью доказательной медицины и достаточно сложным процессом при использовании математических методов для обоснования результатов данных исследований. Особенность медико-биологических исследований заключается в том, что в большинстве случаев они планируются и проводятся не на всей популяции (генеральной совокупности), а на выборке (выборочной совокупности), полученной из этой совокупности, и требуют соблюдения методологических принципов организации научного исследования, обеспечивающих репрезентативность и надежность данных. Следует отметить, что при планировании исследования часто допускаются методологические ошибки, включая некорректное определение объема выборки, несоответствие дизайна, статистических методов анализа целям и задачам исследования и др.

Методологической основой любого медико-биологического исследования является формулировка гипотезы, под которой понимают научное предположение, выдвигаемое для объяснения какого-либо явления и требующее верификации/подтверждения. Гипотеза определяет главное направление научного поиска, является основным методологическим инструментом, организующим весь процесс исследования. Различают научную (исследовательскую) и статистическую гипотезы [1–15]. Научная гипотеза служит основанием для организации исследования и формулируется как предполагаемое решение проблемы или задачи; статистическая – для организации процедуры сравнения регистрируемых признаков (предположение о виде или параметрах распределения случайной величины).

Следует помнить, что теория проверки статистических гипотез является основным инструментом доказательной, а не интуитивной медицины. Научная гипотеза – первична, статистическая – вторична, так как предназначена для оценки результатов, полученных при сравнении показателей состояния организма пациента в норме и при заболевании, до и после лечения,

при применении различных методов лечения; при анализе влияния факторов и т. д. [1–15].

Статистической называют гипотезу о виде неизвестного распределения (например: генеральная совокупность распределена по закону Пуассона или подчиняется закону нормального распределения) или о параметрах известных распределений (средние величины двух совокупностей не равны между собой; дисперсии (разброс значений) двух совокупностей равны между собой и др.).

Статистическая гипотеза – это утверждение относительно неизвестного параметра (случайной величины) генеральной совокупности, которое формулируется для сравнения, оценки надежности связи и которое можно проверить по известным параметрам выборочной совокупности – результатам исследования. В настоящее время существует достаточно много классификаций статистических гипотез, многообразие которых создает сложности при их определении, что обусловлено особенностями проведения исследования

Статистические гипотезы, которые наиболее часто формулируют при проведении медико-биологических исследований, условно классифицируют: по содержанию (гипотезы о законах распределения и гипотезы о параметрах распределения; об однородности двух или нескольких выборок; о числовых значениях характеристик исследуемого явления или процесса), по форме (простые и сложные), по отношению друг к другу (нулевая и альтернативная) и др. Например, гипотеза о том, что случайные величины роста, веса, частоты пульса однородной совокупности людей имеет нормальный закон распределения, является гипотезой о законе распределения. Гипотеза о том, что средние значения артериального давления до и после приема гипотензивного препарата не различаются между собой, является гипотезой о параметрах распределения [1–15].

Основополагающими статистическим гипотезами исследований являются нулевая и альтернативная гипотезы. Проверяемую гипотезу называют нулевой и обозначают H_0 (null hypothesis), которая предполагает отсутствие

статистически значимых различий между анализируемыми параметрами. Ее формулируют в начале исследования для того, чтобы понять, какие доказательства нужно собрать для ее подтверждения или опровержения, какой дизайн исследования предпочесть. Альтернативная гипотеза (alternative hypothesis, H_1) противоположна нулевой, является логическим отрицанием H_0 , предполагает наличие статистически значимых различий между сравниваемыми параметрами. В общепринятом и упрощенном варианте трактования нулевая гипотеза – это гипотеза о сходстве (идентичности), а альтернативная – гипотеза о различии. Например, при изучении «болеутоляющего воздействия» препарата на основании сравнения средних значений индексов болевых ощущений в экспериментальной и контрольных группах, формулирование нулевой гипотезы заключается в следующем: средние значения индекса болевых ощущений в экспериментальной и контрольных группах не отличаются, т. е. препарат не оказывает болеутоляющего действия. Альтернативная гипотеза: средние значения индекса болевых ощущений в экспериментальной и контрольных группах отличаются, т. е. препарат оказывает болеутоляющее действие». Таким образом, при принятии нулевой гипотезы подтверждается сходство воздействия препарата в двух группах (т. е. препарат не оказывает болеутоляющего действия), при ее отклонении подтверждается различие воздействия препарата в двух группах (т. е. препарат оказывает болеутоляющее действие [1–15]).

Простая гипотеза содержит проверку одного предположения (позиции, утверждения) (например, курение является причиной возникновения рака губы). Сложная гипотеза состоит из конечного (не меньше двух) или бесконечного числа простых гипотез (например, сравнение двух методов лечения в зависимости от их влияния на клинические проявления/симптомы заболевания). Интерес исследователя в случае выдвижения сложной гипотезы касается отдельного вывода о влиянии двух методов лечения на каждое клиническое проявление/симптом заболевания, который основан на проверке не одной статистической гипотезы. При проведении клинического испытания нового лекарственного средства, связанного с оценкой его безопасности, также выдвигается несколько гипотез в зависимости от количества по-

бочных эффектов с учетом возраста и пола пациента и т. д. [1–15].

Статистические гипотезы, выдвигаемые исследователями, могут быть направленные (односторонние, one-tail) и ненаправленные (двусторонние, two-tails), которые основаны на определении критической области проверки гипотезы, необходимой для принятия решения об отклонении или не отклонении нулевой гипотезы.

Проверка статистической гипотезы (или процесс проверки статистической значимости) представляет количественную меру аргументов против определенной гипотезы, позволяет обобщить выводы, полученные на выборочной совокупности, и включает следующие этапы [1–15]:

- 1) определение/формулирование нулевой (H_0) и альтернативной гипотезы (H_1) при проведении исследования;

- 2) определение объема выборки, который зависит от задач исследования, степени однородности изучаемого явления и выбранных статистических критериев (тестов);

- 3) выбор требуемого уровня значимости, характеризующего ошибку 1-го типа, ($\alpha = 0,05, 0,01, 0,025, 0,001...$) или вероятности отклонения нулевой гипотезы;

- 4) выбор статистического критерия (теста), соответствующего решаемой задаче (выдвинутой гипотезе);

- 5) вычисление эмпирического значения статистического критерия (теста), на основании которого можно делать вывод об истинности нулевой гипотезы по тесту;

- 6) сравнение вычисленного эмпирического значения статистического критерия с критическим значением (проверка статистики критерия на предмет попадания в критическую область);

- 7) вывод об истинности гипотезы (принятие решения об отклонении нулевой гипотезы). Для большинства статистических критериев (тестов) приемлемо правило: если вычисленное эмпирическое значение статистического критерия (теста), больше, чем критическое, нулевая гипотеза отклоняется;

- 8) интерпретация достигнутого уровня статистической значимости (p -value) результатов исследования.

Логика проверки статистических гипотез заключается в целесообразности и надежно-

сти использования определенного статистического критерия (теста), который включают в себя методы расчета определенного показателя для обоснованного (статистически значимого с позиции доказательной медицины) принятия решения о том, какую из гипотез (нулевую или альтернативную) следует принять, а также ограничения или допустимые условия принятия решения. Полученный в результате проведения медико-биологического исследования называется эмпирическим значением статистического критерия, который сравнивается с известным (например, заданным табличным) эталонным числом, называемым критическим значением критерия. Критические значения приводятся, как правило, для нескольких уровней значимости: 5 % (0,05), 1 % (0,01) или еще более высоких [1–15].

Если полученное исследователем эмпирическое значение статистического критерия оказывается меньше или равно критическому, то нулевая гипотеза не может быть отклонена – считается, что на заданном уровне значимости (то есть при том значении α , для которого рассчитано критическое значение критерия) характеристики распределений совпадают.

Если эмпирическое значение статистического критерия оказывается больше критического, то нулевая гипотеза отвергается и принимается альтернативная гипотеза – характеристики распределений считаются различными с вероятностью безошибочного прогноза $(1 - \alpha)$. Например, если $\alpha = 0,05$ и принята альтернативная гипотеза, то различия статистически значимы с вероятностью безошибочного прогноза 0,95 или 95 %.

Статистическая значимость является одним из основных аспектов проверки гипотезы. Статистическая значимость – максимально допускаемая исследователем вероятность ошибочного отклонения нулевой гипотезы (об отсутствии различия между группами, об отсутствии взаимосвязи признаков и т. п.) на основании результатов применения статистического критерия (теста). Большинство проверяемых гипотез в медико-биологических исследованиях

основаны на сравнении между собой групп людей, которые испытывают влияние различных факторов (например, сравнение эффективности двух методов лечения, обеспечивающих снижение 5-летней смертности пациентов от злокачественных новообразований) [1–15].

В результате проверки гипотезы подтверждается или отклоняется утверждение о том, что наблюдаемые результаты не случайны, а отражают связь между переменными. Особенность доказательной статистики такова, что вы можете доказать альтернативную гипотезу путем опровержения нулевой гипотезы, но вы не доказываете нулевую гипотезу. Нулевая гипотеза либо опровергается в пользу альтернативной, либо нет. Общий подход таков: выдвигается нулевая гипотеза (H_0) о том, что различий в результате применения разных технологий лечения нет (например, 5-летняя смертность от рака молочной железы одинакова в группах с разными методами лечения). Альтернативная гипотеза (H_1) заключается в том, что различия есть, и они не случайны (двухсторонняя альтернативная гипотеза будет означать, что различие эффектов не равно нулю). Выдвигаются две противоречащие друг другу гипотезы: нулевая гипотеза (H_0) о том, что связь (зависимость) случайна, альтернативная (H_1) ей противоречит. Доказав, что гипотеза H_0 несостоятельна, подтверждается, что альтернативная гипотеза верна [1–42].

При проверке статистических гипотез возникают ошибки, которые подразделяют на ошибки 1-го и 2-го типа (таблица 1). Ошибка 1-го типа (α) определяется как вероятность обнаружить различие, тогда как в действительности оно отсутствует («ложноположительный результат»). Другими словами, это вероятность неправомерно отклонить нулевую гипотезу (H_0) в пользу альтернативной (H_1) [1–42].

Уровень статистической значимости (confidence-level – общепринятый термин в международных изданиях, significance level – термин, используемый при переводе в научных журналах на русском языке) – уровень вероятности истинности некоторого результата или гипо-

Таблица 1. Типы ошибок при проверке статистических гипотез

Результаты проверки гипотезы	Истинный, но неизвестный характер взаимодействия	
	гипотеза H_0 верна	гипотеза H_0 не верна (верна H_1)
Принять гипотезу H_0	Правильно (с вероятностью $1-\alpha$)	Ошибка 2 типа (β)
Отвергнуть гипотезу H_0	Ошибка 1 типа (α)	Правильно (с вероятностью $1-\beta$), мощность критерия



Примечание: Если наблюдаемое значение критерия (K) принадлежит критической области ($K_{кр}$ – заштрихованная область на рисунке 1), нулевую гипотезу отвергают, если не принадлежит – не отвергают.

Рисунок 1. Схема распределения критических областей принятия гипотез

тезы, на основе которого принимается решение об отклонении нулевой гипотезы. Другими словами, это допустимая (с точки зрения исследователя) вероятность совершения статистической ошибки 1-го типа – ошибки того, что различия сочтены существенными, а они на самом деле случайны. Обычно используют уровни статистической значимости, равные 0,05, 0,01 и 0,001, которые характеризуют статистические ошибки 1-го типа (α) [1–15]. В настоящее время в связи с развитием компьютерных технологий используется уровень меньше 0,05, если быть точным, то, согласно теории вероятностей и кривой Гаусса, равный 0,0456 (округленно 0,046) и при описании результатов сейчас стали писать то значение, которое дает статистическая программа, и это значение 0,046 стало определяющим ($p = 0,048$ или $p = 0,044$) [1–15].

В медико-биологических исследованиях в качестве критического порога статистической значимости традиционно выбирается уровень 0,05, что допускает наличие ошибки 1-го типа 5 раз в 100 сравнениях. При $p \leq \alpha$ различия принимаются статистически значимыми. И чем меньше p -значение, тем меньше подобных ошибок: например, при $p = 0,01$ считается, что ошибка 1-го типа возможна 1 раз в 100 сравнениях, при $p = 0,001$ – 1 раз в 1000 сравнениях. Однако в пилотных исследованиях допускается уровень значимости $\alpha = 0,1$ для выявления намечающихся различий и/или взаимосвязей с целью дальнейшего планирования на их основе новых исследований с достаточной значимостью [1–42].

Уровень статистической значимости или критические значения критериев определяют при проверке направленных и ненаправленных статистических гипотез, основанных на оценке критической области. Выделяют 3 вида критической области: двухсторонняя,

односторонняя (левосторонняя и правосторонняя) (рисунок 1).

При направленной статистической гипотезе используется односторонний критерий, при ненаправленной гипотезе – двусторонний критерий. Двусторонний критерий более строг, поскольку он проверяет различия в обе стороны, и поэтому то эмпирическое значение критерия, которое ранее соответствовало уровню значимости $p < 0,05$, теперь соответствует лишь уровню $p < 0,10$.

Ненаправленная гипотеза предполагает, что значения переменной в первой выборке отличаются от значений во второй выборке (two-tails), при этом направление различий неизвестно. Направленная гипотеза предполагает, что заранее известно, что значения переменной в первой выборке больше значений во второй (one-tail), или изменения переменной направлены в одну сторону. В научной практике ситуация, описываемая ненаправленной гипотезой, встречается намного чаще, чем соответствующая направленной гипотезе, ибо результат научного поиска, как правило, неизвестен заранее. Это существенно, поскольку большинство программ для статистической обработки вычисляет и предоставляет исследователю величины p для обоих вариантов гипотез (т. н. односторонний, или one-tailed p , и двусторонний, или two-tailed p), при этом величина одностороннего p в два раза меньше величины двустороннего, что создает соблазн использовать при формулировке статистического заключения именно ее. Делать этого не следует, так как данный выбор будет свидетельствовать о нарушении методологии научного исследования [1–15].

Следует также помнить, что при оценке результатов исследования возможно не установление статистических значимых различий, потому что их в действительности нет, либо

из-за отсутствия достаточного количества данных для ее выявления (вследствие слишком малого объема выборки). Вместе с тем, если величина p больше принятого уровня значимости, это не означает, что результат соответствующего статистического критерия (теста) должен непременно игнорироваться при анализе результатов работы. Как уже говорилось, p – это всего лишь степень вероятности нулевой гипотезы, а уровень статистический значимости выбирается произвольно. Проще говоря, если в ходе исследования были получены интересные, логически обоснованные и практически важные результаты, статистическая значимость которых оказалась недостаточной, то их необходимо опубликовать как предварительные (с указанием уровня p), а в дальнейшем – провести еще одно исследование большей мощности, увеличив размер выборки в соответствии с полученными данными [1–15].

Выбор α -уровня неизбежно требует компромисса между значимостью и мощностью, и, следовательно, между вероятностями ошибок 1-го и 2-го типа. Наиболее часто в биомедицинских исследованиях предельно допустимый предел ошибки 1-го типа обычно устанавливается на уровне 5 %, а ошибка 2-го типа – не более 20 % ($\alpha = 0,05$; $\beta \leq 0,2$) [1–42]. При увеличении значения α (например, с 0,01 до 0,05) вероятность ошибки второго типа уменьшается.

Ошибка 2 типа (обозначается β) – это вероятность сделать вывод об отсутствии различия, в то время как фактически оно имеется («ложноотрицательный результат»), т. е. неправомерно принять H_0 . В медико-биологических исследованиях в качестве критического порога принимается значение ошибки 2-го типа $\beta = 0,1$ или $\beta = 0,2$, что в виде статистической мощности ($1 - \beta$), выраженной в процентах, равно 90 % или 80 %. Наиболее чаще выбирается значение статистической мощности, равное 80 %: т. е. вероятность того, что из 100 в 80 случаях действительно существующее различие будет выявлено и в 20 случаях – нет [1–42]. Следует отметить, что последствия этих

длит к риску, иногда неоправданному. Например, если отвергнуть какой-либо эффективный метод лечения (ошибка 1-го типа), то скорее будут применяться другие, менее эффективные. Последствия выбора неправильного лечения (ошибка 2-го типа) могут быть более тяжелыми и непоправимыми (привести к развитию осложнений, инвалидности или к смерти).

При планировании исследования необходимо определить статистическую мощность (statistical power, чувствительность), которая вычисляется как $(1 - \beta)$ и означает вероятность сделать ошибочное заключение об отсутствии различия, в то время как оно имеется на самом деле [1–42].

Статистическая мощность зависит от размера выборки (поскольку связана со стандартной ошибкой, которая в свою очередь уменьшается с увеличением размера выборки), а также от степени различия, которое ожидается обнаружить. Выявление больших различий требует меньшего числа наблюдений и, наоборот, для определения небольших различий потребуется более многочисленная выборка [1–42]. Если планируемая численность выборки не обеспечивает приемлемый уровень статистической мощности (≥ 80 %), чтобы убедительно отвергнуть нулевую гипотезу или согласиться с ней, результаты исследования не будут доказательными. Например, если исследователь хочет определить различие «в среднем весе тела между двумя группами (получавшими и не получавшими препарат, снижающий аппетит) и хочет доказать разницу в 1 кг при стандартном отклонении 10 кг в контрольной и изучаемой группах, то при $\alpha = 0,05$ и мощности 80 % необходимо иметь не менее 1570 людей в каждой группе. Однако, если необходимо оценить различие в 5 кг при стандартном отклонении 10 кг, достаточно включить в группы по 64 человека» [6, 16, 17]. Расчет данного объема выборки для мощности 80 % и двухстороннего критерия уровня значимости $\alpha = 0,05$ осуществлялся с использованием формулы Лера [9]:

$$n = \frac{16}{(\text{стандартизированная разность})^2} = \frac{16}{(5 \setminus 10)^2} = 64,$$

ошибок могут быть неравнозначны, при этом считается, что ошибка 1-го типа ведет к более консервативному или более осторожному решению, а ошибка 2-го типа, наоборот, приво-

где n – число наблюдений, стандартизированная разность (наименьший интересующий эффект) = d (наименьшая разность средних, которая клинически значима/SD (стандартное отклонение).

Кроме быстрых формул (формула Лера), Номограммы Альтмана для расчета объема наблюдений, необходимых для проверки статистической значимости различия двух средних (количественный признак) на уровне значимости 0,05 и при мощности критерия 0,80 для различных значений стандартизированной разности (d/SD); а также для проверки статистической значимости различия частот (качественный признак) двух групп на уровне значимости 0,05 и с мощностью критерия 0,80 можно использовать данные таблиц 2 и 3.

Статистическая мощность особенно важна тогда, когда результаты исследования не являются статистически значимыми. В этом случае исследования с адекватной статистической

мощностью могут интерпретироваться как отрицательный результат: группы в действительности являются схожими, если не эквивалентными. Однако в исследованиях с неадекватной мощностью результаты не являются отрицательными, по ним нельзя сделать никаких заключений.

При проверке гипотез нельзя забывать о так называемой проблеме множественных сравнений, которая заключается в следующем: чем больше статистических гипотез проверяется на одних и тех же данных, тем более вероятно заключение о наличии различий между группами (либо наличии статистической связи признаков), в то время как на самом деле верна нулевая гипотеза об отсутствии различий/

Таблица 2. Пример расчетной таблицы объемов выборки, необходимых для проверки статистической значимости различия двух средних на уровне значимости 0,05 и при мощности критерия 0,80 для различных значений d/SD

d/SD	Расчетный объем выборки для одной группы	Реальный объем выборки для одной группы (N)	Суммарный объем выборки для 2 групп исследования (без учета выбывания)	Суммарный расчетный объем выборки для 2 групп исследования (с учетом 10 % выбывания)	Реальный суммарный объем выборки для 2 групп исследования (с учетом 10 % выбывания)
0,1	1568	1568	3136	3449,6	3450
0,2	392	392	784	862,4	864
0,3	174,2	174	348	382,8	384
0,4	98	98	196	215,6	216
0,5	62,72	63	126	138,6	140
0,6	43,6	44	88	96,8	98
0,7	32	32	64	70,4	72
0,8	24,5	25	50	55	56
0,9	19,4	20	40	44	44
1	15,68	16	32	35,2	36
1,1	13,0	13	26	28,6	30
1,2	10,9	11	22	24,2	26
1,3	9,3	10	20	22	22
1,4	8	8	16	17,6	18
1,5	7,0	7	14	15,4	16

d/SD – минимальное стандартизованное различие (стандартизованная разность); d – наименьшая разность средних, которая клинически значима (для обозначения может использоваться символ « δ », Петри А.); SD – стандартное отклонение (для обозначения может использоваться символ « σ » Петри А.)

Таблица 3. Объемы выборки, необходимые для проверки статистической значимости различия частот двух групп на уровне значимости 0,05 и с мощностью критерия 0,80

Частота группы наблюдения (P_2)	Частота группы наблюдения (P_1)							
	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	200	63	33	21	15	11	8	7
0,2		251	69	33	20	14	10	8
0,3			304	79	36	21	14	10
0,4				351	89	40	23	15
0,5					383	96	43	24
0,6						392	98	44
0,7							374	94
0,8								327

связей [1–21, 16–22, 25, 9, 11, 23, 26]. Многочисленные авторы (Ланг Т. А., Реброва О. Ю., Гржибовский А. М., Наркевич А. Н. и другие) считают, что учет множественных сравнений следует начинать в тех случаях, когда число рассчитываемых значений p превышает 10 [1–21, 16–22, 25, 9, 11, 23, 26]. В таких случаях рекомендуется использовать поправку Бонферрони (вычисление порогового уровня статистической значимости с учетом числа проверенных гипотез), специальные методы множественного сравнения групп (Тьюки, Стьюдента-Ньюмана-Кейсла, Дункана, Шеффе, метод наименьшей значимой разности Фишера (LSD-метод), дисперсионный анализ [1–21, 16–22, 25, 9, 11, 23, 26]. Скорректированная поправка Бонферрони представляет собой установление «нового уровня альфа» с целью определения новой статистической значимости, который определяется по формуле:

$$\text{Новый уровень альфа} = \frac{\text{установленный уровень альфа («исходный», } \alpha = 0,05)}{n(\text{число сравнений исследования})}.$$

Например: В исследовании установлен исходный уровень альфа, равный $\alpha = 0,05$. Необходимо провести 12 сравнений (12 p -значений), следовательно, необходимо рассчитать «новый уровень альфа», на основании которого будут оцениваться статистически значимые различия результатов исследования. «Новый уровень альфа» $= 0,05/12 = 0,004$. Статистически значимыми различиями результатов исследования будут считаться только те результаты, у которых p -значения меньше «нового уровня альфа» ($< 0,004$).

Проверка статистических гипотез выполняется на основе применения статистических критериев (тестов), представляющих собой свод необходимых и достаточных правил, указывающих при каких результатах проверяемая гипотеза принимается или отклоняется. Выбор статистического критерия зависит от его мощности и характеристики данных, включаемых в анализ: типы переменных (количественные, качественные, порядковые), формы распределения и возможные зависимости между ними данных, количества выборочных совокупностей и числа наблюдений в них [1–15].

Статистические критерии, применяемые в биомедицинских исследованиях, подразделяются на параметрические и непараметрические, двусторонние и односторонние. Пара-

метрический критерий – это критерий, применение которого основано на предположении, что закон распределения признаков, характеризующих объектов, известен, в частности, соответствует нормальному закону, и оценка числовых характеристик (параметров) совокупностей значений изучаемых признаков или законов (свойств) их распределения (как правило, нормальном) осуществляется с использованием: t -критерия Стьюдента, критерия F и др. [1–15].

Непараметрический критерий – это критерий, не требующий знания этих параметров и законов распределения, рассчитывается на основе частот и рангов (критерий Вилкоксона, критерий Манна-Уитни, критерий Q Розенбаума и др.), применяется при малом числе наблюдений. Значительная часть статистических критериев носят имена авторов, разработавших теорию и практику их использования,

и обозначаются обычно латинскими или греческими буквами [1–15].

Понятие одностороннего либо двустороннего критерия связано с формулировкой гипотез [10]. Если нулевая гипотеза формулируется о равенстве средних или относительных величин (частота, доля), то для проверки используется двусторонний критерий; если же нулевая гипотеза формулируется о неравенстве – то односторонний [27, 29, 33]. На рисунке 2 показана оценка параметров мощности при проверке гипотезы о равенстве средних ($H_0: \mu_1 = \mu_2$) и равенстве относительных величин ($H_0: P_1 = P_2$) с использованием программы «Statistica».

Важнейшей характеристикой любого статистического критерия является его статистическая мощность – способность выявлять различия, если они есть, или отклонять нулевую гипотезу, если она неверна, характеристика его способности избегать ошибки 2-го типа. Статистическая мощность критерия представляет собой величину $1-\beta$. Мощность может принимать любые значения от 0 до 1 (чем ближе мощность к единице, тем эффективнее критерий). Исследователь должен помнить, что проверка одних и тех же гипотез (задач) может быть решена с помощью разных статистических критериев, в процессе которого обнаруживается, что некоторые критерии позволяют

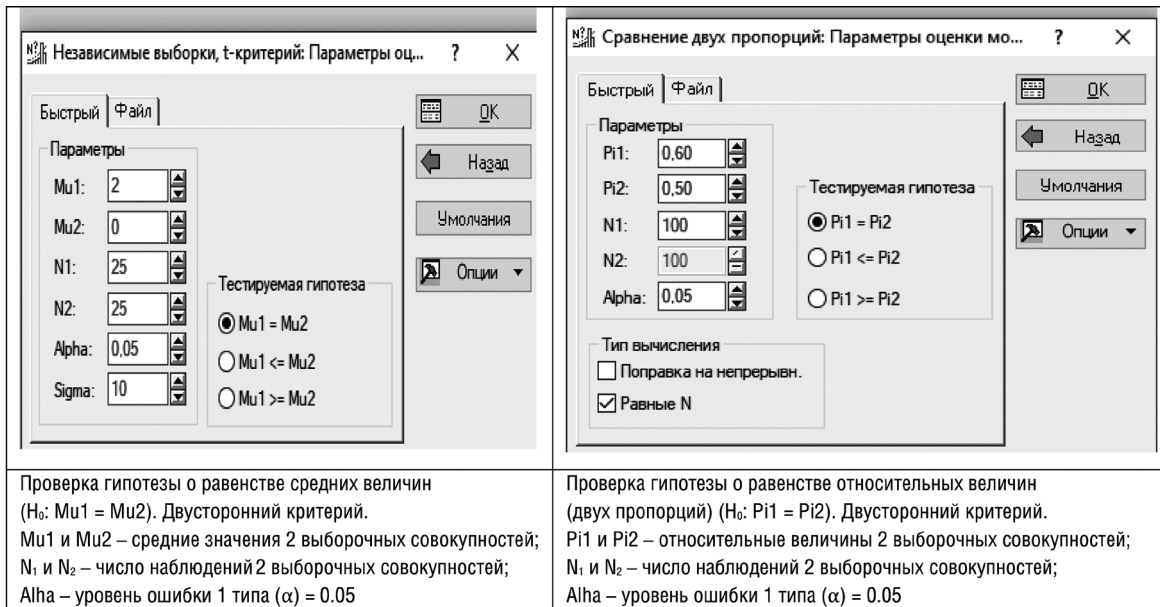


Рисунок 2. Пример оценки параметров мощности при проверке гипотезы о равенстве средних ($H_0: \mu_1 = \mu_2$) и равенстве относительных величин ($H_0: P_1 = P_2$) с использованием программы «Statistica»

выявить различия там, где другие неспособны это сделать, и выбирать статистические критерии с большей мощностью. Параметрические критерии обладают большей мощностью при нормальном распределении случайных величин. В связи с чем для анализа данных выборок, полученных из нормально распределённых генеральных совокупностей, предпочтение следует отдавать параметрическим критериям. Однако, практика показывает, что большинство данных медико-биологических исследований не подчиняются закону нормального распределения, поэтому применение параметрических критериев при анализе результатов может привести к ошибкам статистических выводов. В таких случаях непараметрические критерии оказываются более мощными, т. е. способными с большей статистической значимостью отвергать нулевую гипотезу [1–15].

Для принятия решений о том, какую из гипотез (нулевую или альтернативную) следует принять, используют статистические критерии, которые включают в себя методы расчета определенного показателя (p -значения), на основании которого принимается решение об отклонении или принятии гипотезы, а также правила (условия) принятия решения. Для проверки статистической гипотезы о соответствии эмпирического распределения исследуемой случайной величины определенному теоретическому закону распределения (т. е. формулируется

нулевая гипотеза о том, какому выбранному теоретическому закону распределения подчиняется случайная величина – о соответствии законов распределения, например, закону нормального распределения) используется статистические критерии согласия: χ^2 Пирсона, Колмогорова-Смирнова, Шапиро-Уилка и др. При $p > 0,05$, нулевую гипотезу принимают (анализируемое распределение не отличается от нормального), при $p \leq 0,05$ – отвергают (анализируемое распределение не подчиняется закону нормального распределения), и исследователь должен подобрать для описания переменной более подходящий закон распределения (экспоненциальный, Пуассона и т. п.). Компьютерные программы анализа данных, в т. ч. прикладная программа «Statistica», позволяют реализовать процедуру идентификации закона распределения случайных величин медико-биологического исследования [1–7, 10–12].

Для данных, полученных в метрических шкалах (интервальных или относительных) при распределениях, близких к нормальным, используют параметрические методы, основанные на таких показателях, как среднее и стандартное отклонение (рисунки 3 и 4). В частности, для определения статической значимости различия средних для двух выборок применяют критерий Стьюдента, для трех и с большим числом выборок – F-тест или дисперсионный анализ (ANOVA). Данные, полученные в неметрических (номинативных или порядковых) шкалах,

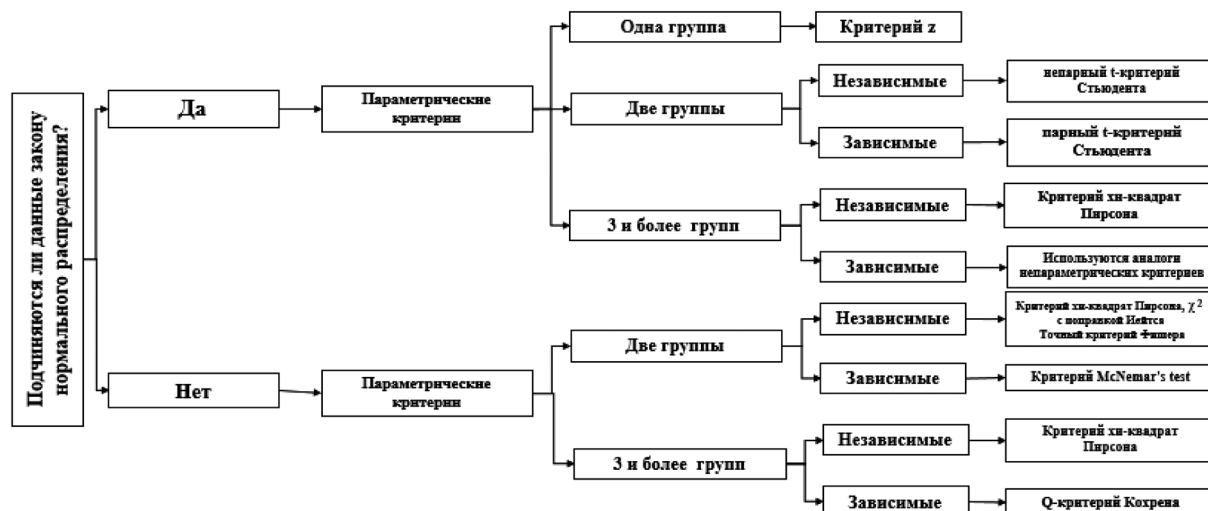


Рисунок 3. Алгоритм проверки статистических гипотез (выбора статистического критерия для анализа качественных данных)

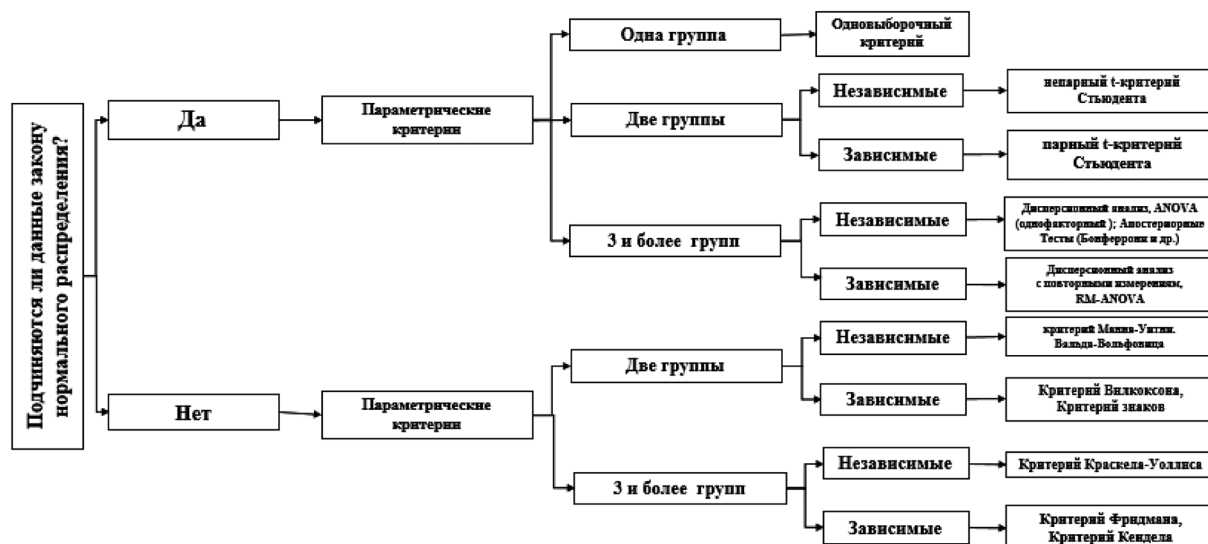


Рисунок 4. Алгоритм проверки статистических гипотез (выбора статистического критерия для анализа количественных данных)

или выборки слишком малые для уверенности в том, что генеральные совокупности, из которых они взяты, подчиняются нормальному распределению, оценивают, используя непараметрические методы – критерий Chi-square (χ^2 , хи-квадрат), Манна-Уитни и др. Выбор статистического метода также зависит от того, являются ли выборки независимыми (т. е. взятыми из двух разных групп исследуемых, используют критерии Манна-Уитни, Kruskal-Wallis) или зависимыми (т. е. отражающими результаты одной и той же группы исследуемых до и после воздействия, используют Вилкоксона, McNemar's test и др.) [1–41].

Кроме того, для проверки статистической гипотезы можно использовать доверительный

интервал (ДИ) – интервал значений признака, рассчитанный по выборке для какого-либо параметра распределения или показателя величины эффекта и с определенной вероятностью (например, с вероятностью безошибочного прогноза 95 %). Ширина ДИ зависит от объема выборки и вариабельности в ней: чем шире ДИ, тем менее точной является выборочная оценка. При увеличении числа наблюдений ДИ сужается, и точность оценки увеличивается [1, 1–22, 25, 26, 33].

Вместе с тем, необходимо отметить различия двух указанных подходов: проверка гипотез позволяет оценить вероятность неверного отклонения нулевой гипотезы (однако ничего не говорит о вероятности альтернативной ги-

потезы), а ДИ позволяет при фиксированном уровне допустимой ошибки оценить размер эффекта и его точность [1, 1–22, 25, 26, 33].

В последнее время для проверки гипотез вместо частотного подхода начинает широко использоваться байесовский подход, который рассматривают как способ переоценки научных представлений с помощью вновь полученных данных. Goodman S. N., описывая метод Байеса указал, что его можно использовать для количественной оценки доказательности данных, основой которой является байесовский критерий, который в простейшей его форме называют также отношением правдоподобия [32]. Минимальный байесовский критерий представляет собой объективный показатель, способный с успехом заменить величину p . В отличие от величины p теоретическое обоснование и интерпретация байесовского критерия позволяют использовать его как в процессе проверки гипотез, так и в процессе принятия решений. Байесовский фактор показывает, что оценка данных с помощью величины p преувеличивает доказательства, опровергающие нулевую гипотезу [16, 9, 32]. И самое главное, байесовский критерий подразумевает включение в анализ прошлого опыта в виде вероятности, что тот или иной вывод правилен либо ошибочен. Байесовский критерий отделяет логическое умозаключение от данных опыта и в то же время дает исследователю возможность комбинировать старую и новую информацию [16, 9, 12].

В настоящее время существует прикладные программы (SPSS, Statistica, SPlus, MedCalc, StatDirect), позволяющие осуществлять проверку статистических гипотез при проведении медико-биологических исследований. Прикладная программа «Statistica» содержит модуль «Анализ мощности», позволяющий провести графические и аналитические процедуры анализа мощности и объема выборки, интервальное оценивание величин, функции распределения, модули, предназначенные для проверки гипотез законов распределения и параметров распределения с использованием различных статистических критериев.

Заключение. Методологической основой проведения научных исследований является формирование статистической гипотезы, подтверждение которой осуществляется на определении статистических критериев. Выбор ста-

тистических критериев проверки статистической гипотезы зависит от дизайна, исследования (цели и задач исследования, типа переменной и распределения данных, количества выборочных совокупностей и числа наблюдений в них).

На этапе планирования исследования важно определить минимальное число наблюдений изучаемой группы (выборочной совокупности) с учетом статистической мощности и ошибок 1-го и 2-го типа, обеспечивающие правомочность результатов проверки гипотез, что позволит оптимизировать использование ресурсов и обосновать статистическую значимость полученных результатов.

Исследователь должен помнить, что бы ни показывали статистические данные, важно уметь аргументировать окончательные результаты и анализировать все данные в совокупности, а не сознательно выбранные для обсуждения. Результаты научных исследований в области медицины следует интерпретировать в контексте характера исследования и других имеющихся данных. При описании результатов исследования необходимо указывать не только величину эффекта, но и точные значения p , а не данные их сравнения с произвольно установленными порогами, так как p -значение является количественной характеристикой только лишь статистической, но не клинической значимости.

Авторы заявляют об отсутствии конфликта интересов.

Литература

1. Военно-медицинская статистика: учебник; под ред. В. И. Кувакина, В. В. Иванова. – СПб.: ВМедА, 2005. – 528 с.
2. Гржибовский, А. М. Выбор статистического критерия для проверки гипотез / А. М. Гржибовский // Экология человека. – 2008. – № 11. – С. 48–57.
3. Гржибовский, А. М., Иванов С. В., Горбатова М. А. Анализ номинальных и ранговых переменных данных с использованием программного обеспечения Statistica и SPSS // Наука и здравоохранение. – 2016. – № 6. – С. 5–39.
4. Жильцов, И. В. Основы медицинской статистики. Дизайн биомедицинских исследований: практическое руководство / И. В. Жильцов, В. М. Семенов, С. К. Зенькова. – Витебск: ВГМУ, 2014. – 154 с.
5. Красько, О. В. Статистический анализ данных в медицинских исследованиях: в 2 ч. / О. В. Красько. – Минск: МГЭУ им. А. Д. Сахарова, 2014. – Ч. I. – 127 с.
6. Купrienko, Н. В. Статистика. Методы анализа распределений. Выборочное наблюдение: учеб. пособие /

- Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – 3-е изд. – СПб.: Изд-во Политехн. ун-та, 2009. – 138 с.
7. Лемешко, Б. Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. – Новосибирск: Изд-во НГТУ, 2011. – 888 с.
8. Медик, В. А. Статистика в медицине и биологии: рук.: в 2 т. / В. А. Медик, М. С. Токмачев, Б. Б. Фишман. – М.: Медицина, 2000. – Т. 1: Теоретическая статистика. – 455 с.
9. Петри, А. Наглядная медицинская статистика / А. Петри, К. Сэбин; пер. с англ. В. П. Леонов. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2019. – 216 с.
10. Реброва, О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ Statistica / О. Ю. Реброва. – М.: МедиаСфера, 2006. – С. 166–176.
11. Реброва, О. Ю. Описание статистического анализа данных в оригинальных статьях. Типичные ошибки / О. Ю. Реброва // Медицинские технологии. – 2011. – № 4. – С. 36–40.
12. Харькова, О. А. Статистические методы и математическое моделирование: учеб. пособие / О. А. Харькова, А. Г. Соловьев. – Архангельск: Изд-во Северного государственного медицинского университета, 2017. – 164 с.
13. Шеламова, М. А. Статистический анализ медико-биологических данных с использованием программы Excel: учеб.-метод. пособие / М. А. Шеламова, Н. И. Инсарова, В. Г. Лещенко. – Минск: БГМУ, 2010. – 96 с.
14. Lehmann, E. L. The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? // J Am Stat Assoc. – 1993. – Vol. 88. – P. 1242–1249.
15. Moher, D., Dulberg C. S., Wells G. A. Statistical power, sample size, and their reporting in randomized controlled trials // JAMA. – 1994. – Vol. 272. – P. 122–124.
16. Ланг, Т. А., Сесик М. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов / Т. А. Ланг, М. Сесик; пер. с англ. под ред. В. П. Леонова. – М.: Практическая медицина, 2011. – 480 с.
17. Ласта, Джон М., Власов В. В., Вербицкая Е. В. Эпидемиологический словарь. – М., 2009. – 316 с.
18. Банержи, А. Медицинская статистика понятным языком: вводный курс / А. Банержи. – М.: Практическая медицина, 2007. – 287 с.
19. Власов, В. В. Значение научных публикаций в специализированных журналах / В. В. Власов // Росс. вестник акушера-гинеколога. – 2010. – № 10(4). – С. 72–75.
20. Гланц, С. Медико-биологическая статистика / С. Гланц. – М.: Практика, 1999. – 334 с.
21. Гринхальх, Т. Основы доказательной медицины / Т. Гринхальх. – М.: ГЭОТАР-МЕД, 2004. – 240 с.
22. Леонов, В. П. Ошибки статистического анализа биомедицинских данных / В. П. Леонов // Международный журнал медицинской практики. – 2007. – № 2. – С. 19–35.
23. Флетчер, Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. – М.: Медиа Сфера, 2004. – 297 с.
24. Мороз, И. Н. Оценка статистической и клинической значимости в медико-биологических исследованиях / И. Н. Мороз // Медицинский журнал. – 2020. – № 1(71). – С. 97–103.
25. Павлова, В. Ю. Основные вопросы статистического анализа в медицинских исследованиях / В. Ю. Павлова // Клиническая онкогематология. – 2009. – Т. 2, № 4. – С. 374–377.
26. Altman, D. G. Statistical guidelines for contributors to medical journals / D. G. Altman, S. M. Gore, M. J. Gardner, S. J. Pocock // British Medical Journal. – 1983. – Vol. 286. – P. 1489–1493.
27. Berger, J. O., Sellke T. Testing a point null hypothesis: the irreconcilability of P values and evidence // J Am Stat Assoc. – 1987. – Vol. 82. – P. 112–122.
28. Brophy, J. M., Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes // JAMA. – 1995. – Vol. 273. – P. 871–875.
29. Burton, P. R., Gurrin L. C., Campbell M. J. Clinical significance not statistical significance: a simple Bayesian alternative to p-values // J Epidemiol Community Health. – 1998. – Vol. 52. – P. 318–323.
30. Gardner, M. J, Altman. D. G. Confidence intervals rather than P values: estimation rather than hypothesis testing / M. J. Gardner, D. G. Altman // BMJ. – 1986. – Vol. 292. – P. 746–750.
31. Goodman, S. N. Toward evidence-based medical statistics. 1: The P value fallacy // Ann Intern Med. – 1999. – Vol. 130. – P. 995–1004.
32. Goodman, S. N. Towards evidence-based medical statistics: 1: The Bayes factor // Ann Intern Med. – 1999. – Vol. 130. – P. 1005–1013.
33. Feinstein, A. R. P-values and confidence intervals: two sides of the same unsatisfactory coin // J Clin Epidemiol. – 1998. – Vol. 51. – P. 355–360.
34. Fisher, R. A. Statistical methods and scientific inference. – London: Collins Macmillan, 1973.
35. Freiman, J. A., Chalmers T. C., Smith H., Kuebler R. R. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. In: Bailar J. C., Mosteller F. eds. Medical uses of statistics. – Boston, Ma: NEJM Books, 1992. – P. 357–373.
36. Easterbrook, P. J., Berlin J. A., Gopalan R., Matthews D. R. Publication bias in clinical research // Lancet. – 1991. – Vol. 337. – P. 867–72.
37. Kline, R. B. Beyond significance testing: Performing data analysis methods in behavioral research. – APA, 2004.
38. Lang, T. Twenty Statistical Errors Even YOU Can Find in Biomedical Research Articles // Croatian Medical Journal. – 2004. – Vol. 45(4). – P. 361–370.
39. Mulward, S., Gwutzsche P. C. Sample size of randomized double-blind trials 1976–1991 // Dan Med Bull. – 1996. – Vol. 13. – P. 96–98.
40. Sterne, J. A. C. Sifting the evidence – what's wrong with significance tests? / J. A. C. Sterne, G. D. Smith // British Medical Journal. – 2001. – Vol. 322. – P. 226–31.
41. Thompson, B. "Statistical", "practical" and "clinical": How many kinds of significance do counselors need to consider? // Journal of Counseling and Development. – 2002. – Vol. 80, № 1. – P. 64–71.
42. Yancy, J. M. Ten rules for reading clinical research reports / J. M. Yancey // Am J Surg. – 1990. – Vol. 159(6). – P. 533–9.

43. Нелесова, Д. К. Проверка статистических гипотез: теория и практика в современных исследованиях / Д. К. Нелесова // *Ео ipso*. – 2025. – № 3. – С. 7–9.

44. Джумаева, А. Различные виды статистических гипотез и их применения / А. Джумаева // *А Posteriori*. – 2025. – № 3. – С. 11–13.

45. Шапиро, В. Я. Системный анализ. Проверка статистических гипотез: учеб. пособие / В. Я. Шапиро, Т. А. Осечкина, С. И. Затенко. – СПб.: ООО «Медиапапир», 2025. – 58 с. – ISBN 978-5-00110-519-0.

References

1. *Voenno-medicinskaya statistika: uchebnik* [Military medical statistics: a textbook]; pod red. V. I. Kuvakina, V. V. Ivanova. – SPb.: VMedA, 2005. – 528 p. [in Russian].

2. Grzhibovskij, A. M. Vybor statisticheskogo kriteriya dlya proverki gipotez [Select statistical criterion for hypothesis testing] / A. M. Grzhibovskij // *Ekologiya cheloveka*. – 2008. – № 11. – С. 48–57 [in Russian].

3. Grzhibovskij, A. M., Ivanov S. V., Gorbatoва M. A. Analiz nominal'nyh i rangovyh peremennyh dannyh s ispol'zovaniem programmnogo obespecheniya Statistica i SPSS [Analysis of nominal and ranking variables using the software Statistica and SPSS] // *Nauka i zdavoohranenie*. – 2016. – № 6. – С. 5–39 [in Russian].

4. Zhil'cov, I. V. Osnovy medicinskoj statistiki. Dizajn biomedicinskih issledovaniy: prakticheskoe rukovodstvo [Basics of health statistics. Design of biomedical research: practical guide] / I. V. Zhil'cov, V. M. Semenov, S. K. Zen'kova. – Vitebsk: VGMU, 2014. – 154 p. [in Russian].

5. Kras'ko, O. V. Statisticheskij analiz dannyh v medicinskih issledovaniyah: v 2 ch. [Statistical analysis of data in medical research: in 2 parts] / O. V. Kras'ko. – Minsk: MGEU im. A. D. Saharova, 2014. – Ch. I. – 127 p. [in Russian].

6. Kuprienko, N. V. Statistika. Metody analiza raspredelenij. Vyborochnoe nablyudenie: ucheb. posobie [Statistics. Methods of analysis of distributions. Random observation. 3rd ed.: study. manual.] / N. V. Kuprienko, O. A. Ponomareva, D. V. Tihonov. – 3-e izd. – SPb.: Izd-vo Politekh. un-ta, 2009. – 138 p. [in Russian].

7. Lemeshko, B. Yu. Statisticheskij analiz dannyh, modelirovanie i isledovanie veroyatnostnyh zakonornostej. Komp'yuternyj podhod: monografiya [Statistical analysis of data, modelling and probabilistic analysis. Computer approach: monograph] / B. Yu. Lemeshko, S. B. Lemeshko, S. N. Postovalov, E. V. Chimitova. – Novosibirsk: Izd-vo NGTU, 2011. – 888 p. [in Russian].

8. Medik, V. A. Statistika v medicine i biologii: ruk. v 2 t. T. 1. Teoreticheskaya statistika [Statistics in medicine and biology: hands. in 2 vol. Vol. 1. Theoretical statistics] / V. A. Medik, M. S. Tokmachev, B. B. Fishman. – M.: Medicina, 2000. – 455 p. [in Russian].

9. Petri, A. Naglyadnaya medicinskaya statistika [Visual medical statistics] / A. Petri, K. Sebin; per. s angl. V. P. Leonov. – 3-e izd., pererab. i dop. – M.: GEOTAR-Media, 2019. – 216 p. [in Russian].

10. Rebrova, O. Yu. Statisticheskij analiz medicinskih dannyh. Primenenie paketa prikladnyh programm Statistica [Statistical analysis of medical data. Application of the Statistica software package] / O. Yu. Rebrova. – M.: MediaSfera, 2006. – С. 166–176 [in Russian].

11. Rebrova, O. Yu. Opisaniye statisticheskogo analiza dannyh v original'nyh stat'yah. Tipichnye oshibki [Description of statistical analysis in original articles. Common errors] / O. Yu. Rebrova // *Medicinskie tekhnologii*. – 2011. – № 4. – С. 36–40 [in Russian].

12. Har'kova, O. A., Solov'ev A. G. Statisticheskie metody i matematicheskoe modelirovanie: uchebnoe posobie [Statistical methods and mathematical modelling: a training manual] / O. A. Har'kova, A. G. Solov'ev. – Arhangel'sk: Izd-vo Severnogo gosudarstvennogo medicinskogo universiteta, 2017. – 164 p. [in Russian].

13. Shelamova, M. A. Statisticheskij analiz mediko-biologicheskikh dannyh s ispol'zovaniem programmy Excel: ucheb.-metod. posobie [Statistical analysis of medical and biological data using the program Excel: study-method. manual] / M. A. Shelamova, N. I. Insarova, V. G. Leshchenko. – Minsk: BGMU, 2010. – 96 p. [in Russian].

14. Lehmann, E. L. The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? // *J Am Stat Assoc*. – 1993. – Vol. 88. – P. 1242–1249.

15. Moher, D., Dulberg C. S., Wells G. A. Statistical power, sample size, and their reporting in randomized controlled trials // *JAMA*. – 1994. – Vol. 272. – P. 122–124.

16. Lang, T. A., Sesik M. Kak opisyvat' statistiku v medicine. Annotirovannoe rukovodstvo dlya avtorov, redaktorov i recenzentov [How to describe statistics in medicine. Annotated guide for authors, editors and reviewers] // T. A. Lang, M. Sesik; per. s angl. pod red. V. P. Leonova. – M.: Prakticheskaya medicina, 2011. – 480 p. [in Russian].

17. Lasta, Dzhon M., Vlasov V. V., Verbickaya E. V. Epidemiologicheskij slovar' [Epidemiological dictionary]. – M., 2009. – 316 p. [in Russian].

18. Banerzhi, A. Medicinskaya statistika ponyatnym yazykom: vvodnyj kurs [Medical statistics in clear language: introductory course] / A. Banerzhi. – M.: Prakticheskaya medicina, 2007. – 287 p. [in Russian].

19. Vlasov, V. V. Znachenie nauchnyh publikacij v specializirovannyh zhurnalakh [Importance of scientific publications in specialized journals] / V. V. Vlasov // *Ross. vestnik akushera-ginekologa*. – 2010. – № 10(4). – С. 72–75 [in Russian].

20. Glanc, S. Mediko-biologicheskaya statistika [Medico-biological statistics] / S. Glanc. – M.: Praktika, 1999. – 334 p. [in Russian].

21. Grinhal'h, T. Osnovy dokazatel'noj mediciny [Basics of evidence-based medicine] / T. Grinhal'h. – M.: GEOTAR-MED, 2004. – 240 p. [in Russian].

22. Leonov, V. P. Oshibki statisticheskogo analiza biomedicinskih dannyh [Errors in the statistical analysis of biomedical data] / V. P. Leonov // *Mezhdunarodnyj zhurnal medicinskoj praktiki*. – 2007. – № 2. – С. 19–35. – <http://www.biometrika.tomsk.ru/error.htm> [in Russian].

23. Fletcher, R., Fletcher S., Vagner E. Klinicheskaya epidemiologiya. Osnovy dokazatel'noj mediciny [Clinical epidemiology. Fundamentals of evidence-based medicine]. – M.: Media Sfera, 2004. – 297 p. [in Russian].

24. Moroz, I. N. Ocenka statisticheskoy i klinicheskoy znachimosti v mediko-biologicheskikh issledovaniyah [Evaluation of statistical and clinical significance in medical and biological research] / I. N. Moroz // *Medicinskij zhurnal*. – 2020. – № 1(71). – С. 97–103 [in Russian].

25. Pavlova, V. Yu. Osnovnye voprosy statisticheskogo analiza v medicinskih issledovaniyah [Main issues of sta-

- tistical analysis in medical research] / V. Yu. Pavlova // Klinicheskaya onkogematologiya. – 2009. – Vol. 2, № 4. – S. 374–377.
26. Altman, D. G. Statistical guidelines for contributors to medical journals / D. G. Altman, S. M. Gore, M. J. Gardner, S. J. Pocock // British Medical Journal. – 1983. – Vol. 286. – P. 1489–1493.
27. Berger, J. O., Sellke T. Testing a point null hypothesis: the irreconcilability of P values and evidence // J Am Stat Assoc. – 1987. – Vol. 82. – P. 112–122.
28. Brophy, J. M., Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes // JAMA. – 1995. – Vol. 273. – P. 871–875.
29. Burton, P. R., Gurrin L. C., Campbell M. J. Clinical significance not statistical significance: a simple Bayesian alternative to p-values // J Epidemiol Community Health. – 1998. – Vol. 52. – P. 318–323.
30. Gardner, M. J., Altman. D. G. Confidence intervals rather than P values: estimation rather than hypothesis testing / M. J. Gardner, D. G. Altman // BMJ. – 1986. – Vol. 292. – P. 746–50.
31. Goodman, S. N. Toward evidence-based medical statistics. 1: The P value fallacy // Ann Intern Med. – 1999. – Vol. 130. – P. 995–1004.
32. Goodman, S. N. Towards evidence-based medical statistics: 1: The Bayes factor // Ann Intern Med. – 1999. – Vol. 130. – P. 1005–1013.
33. Feinstein, A. R. P-values and confidence intervals: two sides of the same unsatisfactory coin // J Clin Epidemiol. – 1998. – Vol. 51. – P. 355–360.
34. Fisher, R. A. Statistical methods and scientific inference. – London: Collins Macmillan, 1973.
35. Freiman, J. A., Chalmers T. C., Smith H., Kuebler R. R. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. In: Bailar J. C., Mosteller F. eds. Medical uses of statistics. – Boston, Ma: NEJM Books. – 1992. – P. 357–373.
36. Easterbrook, P. J., Berlin J. A., Gopalan R., Matthews D. R. Publication bias in clinical research // Lancet. – 1991. – Vol. 337. – P. 867–72.
37. Kline, R. B. Beyond significance testing: Performing data analysis methods in behavioral research. – APA, 2004.
38. Lang, T. Twenty Statistical Errors Even YOU Can Find in Biomedical Research Articles // Croatian Medical Journal. – 2004. – Vol. 45(4). – P. 361–370.
39. Mulward, S., Gshtzsche P. C. Sample size of randomized double-blind trials 1976–1991 // Dan Med Bull. – 1996. – Vol. 13. – P. 96–98.
40. Sterne, J. A. C. Sifting the evidence – what’s wrong with significance tests? / J. A. C. Sterne, G. D. Smith // British Medical Journal. – 2001. – Vol. 322. – P. 226–231.
41. Thompson, B. “Statistical”, “practical” and “clinical”: How many kinds of significance do counselors need to consider? // Journal of Counseling and Development. – 1990. – Vol. 159(6). – P. 533–9.
42. Yancy, J. M. Ten rules for reading clinical research reports / J. M. Yancey // Am J Surg. – 1990. – Vol. 159(6). – P. 533–9.
43. Nepesova, D. K. Proverka statisticheskikh gipotez: teoriya i praktika v sovremennykh issledovaniyakh [Verification of statistical hypotheses: theory and practice in modern research] / D. K. Nepesova // Eo ipso. – 2025. – № 3. – S. 7–9.
44. Dzhumaeva, A. Razlichnye vidy statisticheskikh gipotez i ih primeneniya [Different types of statistical hypotheses and their applications] / A. Dzhumaeva // A Posteriori. – 2025. – № 3. – S. 11–13 [in Russian].
45. Shapiro, V. Ya. Sistemnyy analiz. Proverka statisticheskikh gipotez: uchebnoe posobie [System analysis. Statistical hypothesis testing: tutorial] / V. Ya. Shapiro, T. A. Osechkina, S. I. Zatenko. – Sankt-Peterburg: OOO “Mediapapir”, 2025. – 58 p. [in Russian].

Поступила 08.09.2025 г.