

ISSN 2225-6016

# ВЕСТНИК

*Смоленской государственной  
медицинской академии*

*Том 18, №4*

2019



**СПОСОБ ВЫЧИСЛЕНИЯ МОЩНОСТИ  $t$ -КРИТЕРИЯ В ФАРМАКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**

© Лямец Л.Л., Евсеев А.В., Козлова Е.К., Колпакова М.А.

*Смоленский государственный медицинский университет, Россия, 214019, Смоленск, ул. Крупской, 28**Резюме*

**Цель.** В статистическом анализе экспериментальных данных существует ряд методов проверки статистических гипотез, основанных на распределении Стьюдента. Эти методы имеют в своей основе алгоритм вычислений, имеющий общее популярное название –  $t$ -критерий. Наиболее часто  $t$ -критерий применяется для проверки статистической гипотезы, предполагающей отсутствие значимых различий между средними значениями двух выборок, распределенных по нормальному закону. Описанная в большинстве литературных источников процедура принятия решения в отношении проверяемой статистической гипотезы при использовании  $t$ -критерия основана на вычислении  $p$ -величины и ее последующем сравнении с зафиксированным уровнем значимости  $\alpha$ , который определяет вероятность ошибки первого рода и специфичность критерия  $1-\alpha$ . При этом математические вычисления, необходимые для количественной оценки ошибки второго рода и чувствительности, как правило, не приводятся. Цель теоретического исследования заключалась в разработке удобного для практического использования способа вычисления мощности и ошибки второго рода для  $t$ -критерия. Способ предназначается для применения в медико-биологических исследованиях, а также для выявления значимых различий в состояниях, вызванных фармакологическим воздействием.

**Методика.** Проведен обзорный анализ печатных и электронных публикаций по применению  $t$ -критерия для проверки статистической гипотезы об отсутствии значимых различий между средними значениями двух выборок, распределенных по нормальному закону. Описаны особенности выполнения вычислительных процедур, связанных с реализацией  $t$ -критерия, а также вычислением  $p$ -величины, ошибки второго рода и мощности  $t$ -критерия на основе интегрирования распределения Стьюдента и нецентрального распределения Стьюдента. Выявлены основные вычислительные действия, которые могут представлять определенные трудности для исследователей, не имеющих специальной математической подготовки. Рассмотрена возможность автоматизации сложных вычислительных процедур с использованием доступного программного обеспечения и информационных ресурсов сети Интернет. Вычисления мощности и ошибки второго рода приведены для двух вариантов  $t$ -критерия. В первом варианте дисперсии статистических совокупностей предполагаются неизвестными и одинаковыми. Во втором случае дисперсии предполагаются неизвестными и различающимися между собой.

**Результаты.** Разработан способ вычисления ошибки второго рода и мощности  $t$ -критерия с использованием современных информационных технологий, находящихся в открытом доступе. В основе способа лежит автоматизация операций интегрирования распределения Стьюдента а также операций интегрирования нецентрального распределения Стьюдента. Автоматизация вычислений существенно упрощает вычисление величин, необходимых для проверки гипотезы об отсутствии значимых различий между средними значениями двух выборок, распределенных по нормальному закону, а также вычисления ошибки второго рода и мощности  $t$ -критерия. Разработанный способ существенно снижает временные затраты на выполнение требуемых математических вычислений и уменьшает вероятность вычислительных ошибок за счет автоматизации производимых вычислений.

**Заключение.** В результате теоретического исследования разработан способ вычисления мощности и ошибки второго рода для  $t$ -критерия при его использовании для проверки статистической гипотезы об отсутствии значимых различий между средними значениями двух выборок, распределенных по нормальному закону. Предложенный способ позволяет автоматизировать сложные вычислительные операции, связанные с интегрированием функций плотности распределения вероятностей. Для автоматизации вычислений предложены встроенные статистические функции табличного процессора Excel и интернет-ресурса Keisan Online Calculator, находящегося в открытом доступе. Предложенный способ вычисления мощности и ошибки второго рода для  $t$ -критерия может представлять практический интерес для научных работников, осуществляющих исследования в области фармакологии и доказательной медицины и использующих в своей работе статистические методы анализа экспериментальных данных.

*Ключевые слова:* мощность статистического критерия, ошибки первого и второго рода,  $t$ -распределение Стьюдента, нецентральное  $t$ -распределение Стьюдента,  $t$ -критерий

METHOD FOR CALCULATING THE POWER OF THE  $t$ -TEST IN MEDICAL AND PHARMACOLOGICAL STUDIES

Lyamec L.L., Evseev A.V., Kozlova E.K., Kolpakova M.A.

*Smolensk State Medical University, 28, Krupskoj St., 214019, Smolensk, Russia**Abstract*

**Objective.** In statistical analysis of experimental data, there is a number of methods to test statistical hypotheses based on Student's distribution. These methods are based on the algorithm of calculations, which has a common popular name  $t$ -test. The most common use of the  $t$ -test is to test the statistical hypothesis that assumes that there are no significant differences between the mean values of two samples distributed according to the normal law. The procedure of decision-making in relation to the tested statistical hypothesis described in most literature sources when using the  $t$ -test is based on the calculation of the  $p$ -value and its subsequent comparison with the fixed level of significance  $\alpha$ , which determines the probability of error of the first kind and the specificity of the criterion  $1-\alpha$ . In this case, the mathematical calculations necessary to quantify the error of the second kind and sensitivity are not given, as a rule. The purpose of the theoretical study was to develop a convenient for practical use method for calculating the power and error of the second kind for the  $t$ -test. The method is intended for use in biomedical research, as well as to identify significant differences in conditions caused by pharmacological effects.

**Method.** A review analysis of printed and electronic publications on the application of the  $t$ -test to test the statistical hypothesis that there are no significant differences between the mean values of the two samples distributed according to the normal law was carried out. The performance of computational procedures associated with the implementation of  $t$ -test and calculating  $p$ -values, error of the second kind and power of  $t$ -test are described, based on the integration of the Student's distribution and non-central Student's distribution. The basic computational actions which can present certain difficulties for researchers who do not have basic physical and mathematical education are revealed. The possibility of automation of complex computational procedures using available software and information resources of the Internet is considered. Practical examples of computing power and errors of the second kind for the  $t$ -test are given for the case when the variances of statistical aggregates are assumed to be unknown and the same, as well as for the case when the variances are assumed to be unknown and different from each other.

**Results.** A method for calculating the error of the second kind and the power of the  $t$ -test using modern information technologies in the public domain is developed. The method is based on automation of Student's distribution integration operations as well as Student's non-central distribution integration operations. Automation of calculations significantly simplifies the calculation of the values necessary to test the hypothesis of no significant differences between the mean values of two samples distributed according to the normal law, as well as the calculation of the error of the second kind and the power of the  $t$ -test. The developed method significantly reduces the time spent on performing the required mathematical calculations and reduces the probability of computational errors by automating the calculations.

**Conclusion.** As a result of the theoretical study, a method for calculating the power and error of the second kind for the  $t$ -test is developed to be used to test the statistical hypothesis that there are no significant differences between the mean values of two samples distributed according to the normal law. The proposed method makes it possible to automate complex computational operations associated with the integration of probability distribution density functions. For automation of calculations the built-in statistical functions of the table processor Excel and the Internet resource Keisan Online Calculator, which is in open access are offered. The proposed method of calculating the power and error of the second kind for the  $t$ -test may be of practical interest for scientists conducting research in the field of pharmacology and evidence-based medicine and using statistical methods of analysis of experimental data.

**Keywords:** power of statistical criterion, errors of the first and second kind, Student's  $t$ -distribution, Student's non-central  $t$ -distribution,  $t$ -test

## Введение

Выявление и изучение взаимосвязей между процессами и явлениями в живом организме представляет определенный практический и научный интерес для фармакологии и медицины. Для количественного описания процессов и явлений можно использовать физические параметры (признаки) – непрерывные физические величины, измеряемые в шкале интервалов или шкале отношений при помощи специальных приборов. Формализованное описание зависимостей между признаками при помощи математических конструктов является одной из форм представления научных знаний. Рассмотрим исследовательскую задачу, которая состоит в обосновании каузальной зависимости между признаками  $X$  и  $Y$ , при этом предполагается, что признак  $X$  является факторным, т.е. количественно характеризует воздействие, а результирующий признак  $Y$  количественно характеризует реакцию (отклик). При планировании наиболее простого фармакологического или медико-биологического эксперимента факторный признак  $X$  может иметь два уровня  $X_1$  и  $X_2$ , где каждый уровень представляет собой строго определенное, контролируемое воздействие. Для проведения эксперимента из исследуемой и формально заданной генеральной совокупности на основе специальных методов, обеспечивающих случайность отбора и репрезентативность, формируются выборочные совокупности  $V_1$  и  $V_2$ , содержащие  $N_1$  и  $N_2$  единиц наблюдения соответственно. На совокупность  $V_1$  оказывается факторное воздействие  $X_1$ , а на совокупность  $V_2$  – факторное воздействие  $X_2$ . Признак  $Y$  количественно выражает исследуемое состояние, вызванное факторным признаком. В исследуемых выборочных совокупностях  $V_1$  и  $V_2$  признак  $Y$  измеряется у каждой единицы наблюдения. В результате проведенных измерений формируются два выборочных множества числовых значений  $Y_1$  и  $Y_2$ . Множество  $Y_1$  содержит выборочные числовые значения результирующего признака, полученные в результате воздействия фактора с уровнем  $X_1$  на единицы наблюдения выборочной совокупности  $V_1$ . Соответственно множество  $Y_2$  содержит выборочные числовые значения результирующего признака, полученные в результате воздействия фактора с уровнем  $X_2$  на единицы наблюдения выборочной совокупности  $V_2$ . Следовательно, выборочные числовые множества  $Y_1$  и  $Y_2$  (далее – выборки) количественно выражают состояние единиц наблюдения, вызванные факторными воздействиями  $X_1$  и  $X_2$ . Если при помощи статистических методов анализа будет обосновано значимое различие между выборками  $Y_1$  и  $Y_2$ , то этот факт дает основание утверждать о влиянии факторного признака  $X$  на результирующий признак  $Y$ .

В медико-биологических исследованиях, а также в исследованиях, имеющих своей целью выявление значимых различий в состояниях, вызванных фармакологическим воздействием, широко применяются методы параметрического статистического анализа. Наиболее распространенным из них является метод сравнения средних значений анализируемых выборок. По своей сути среднее значение является количественной характеристикой типичности. Для физической измеряемого признака, т.е. измеряемого в шкале интервалов или шкале отношений, типичное – это наиболее вероятное значение, в окрестности которого сосредотачиваются остальные значения исследуемого признака. Среднее значение признака является адекватной оценкой типичности в том случае, если исследуемый признак в статистической совокупности распределен по нормальному закону. Поэтому аналитическим действиям со средними значениями исследуемых признаков должна предшествовать проверка статистической гипотезы о соответствии эмпирического распределения теоретическому нормальному распределению, в котором одним из параметров является среднее арифметическое значение [1].

Если можно обоснованно полагать, что анализируемые выборки  $Y_1$  и  $Y_2$  распределены по нормальному закону, то результат сравнения средних значений  $\bar{Y}_1$  и  $\bar{Y}_2$  этих выборок позволяет сделать заключение об однородности этих совокупностей. Если средние значения  $\bar{Y}_1$  и  $\bar{Y}_2$  значимо различаются, то исследуемые статистические совокупности неоднородны и значимо различаются, поскольку различаются типичностью проявления изучаемого результирующего признака.

Сравнение средних значений  $\bar{Y}_1$  и  $\bar{Y}_2$  двух нормально распределенных выборок  $Y_1$  и  $Y_2$  представляет собой формальную процедуру, которая называется  $t$ -критерием. Алгоритм этого критерия направлен на проверку нулевой статистической гипотезы  $H_0$  о том, что выборочные средние значения значимо не различаются, т.е. наблюдаемые различия не выходят за пределы статистической ошибки. В качестве альтернативной статистической гипотезы  $H_1$  можно рассматривать одностороннюю гипотезу о том, что среднее значение одной выборки значимо больше, чем среднее значение другой, т.е. наблюдаемые различия выходят за пределы статистической ошибки. Поскольку исследуемые выборки  $Y_1$  и  $Y_2$  являются случайными продуктами, а выборочные средние  $\bar{Y}_1$  и  $\bar{Y}_2$  – случайными величинами, то выводы, сделанные по

результатам сравнения средних значений, могут носить только вероятностный характер.

Популярное изложение  $t$ -критерия в доступных литературных источниках ограничивается нахождением  $p$ -величины, выражающей вероятность гипотезы  $H_0$ , или вычислением критического значения статистики Стьюдента. Эти величины необходимы для проверки гипотезы  $H_0$  при зафиксированной вероятности ошибки первого рода  $\alpha$  и специфичности  $1-\alpha$ . Вопросы о вычислении мощности критерия и вероятности ошибки второго рода обычно не рассматриваются. В отдельных публикациях эти вопросы излагаются на уровне общих теоретических положений без примеров конкретных вычислений.

Отсутствие в публикациях информации о мощности  $t$ -критерия и вероятности ошибки второго рода не дает полного представления о состоятельности статистических выводов и снижает качество результатов исследования в целом. Из вышесказанного следует, что разработка способа, позволяющего автоматизировать вычисления мощности  $t$ -критерия и вероятности ошибки второго рода, является актуальной задачей.

Цель исследования заключалась в разработке удобного для практического использования способа вычисления мощности и ошибки второго рода для  $t$ -критерия. Способ предназначен для применения в медико-биологических исследованиях, а также для выявления значимых различий в состояниях, вызванных фармакологическим воздействием.

## Методика

В большинстве научных публикаций по фармакологическим и медико-биологическим исследованиям применение  $t$ -критерия происходит без вычисления его мощности. В статистических выводах о результатах проверки статистической гипотезы  $H_0$ , сделанных при заданной вероятности ошибки первого рода  $\alpha$  и специфичности  $1-\alpha$ , не указывается вероятность ошибки второго рода. Разработанный способ вычисления мощности  $t$ -критерия следует рассматривать как неотъемлемую часть этого критерия. Поэтому следует представить алгоритм расчета для  $t$ -критерия совместно с разработанным способом вычисления мощности и вероятности ошибки второго рода.

Рассмотрим применение  $t$ -критерия и способа оценки его мощности для анализа результатов фармакологического исследования, целью которого является обоснование влияния факторного признака  $X$  на результативный признак  $Y$ . В проводимых экспериментах признак  $X$  имеет два строго определенных и контролируемых уровня  $X_1$  и  $X_2$ . Пусть в результате выборочных исследований сформированы два множества числовых значений  $Y_1$  и  $Y_2$ , содержащие  $N_1$  и  $N_2$  элементов соответственно. Также будем полагать, что выборки  $Y_1$  и  $Y_2$  являются малыми, т.е. их объем не превосходит 30 значений. Выборочные числовые множества  $Y_1$  и  $Y_2$  количественно выражают состояние единиц наблюдения, вызванные факторными воздействиями  $X_1$  и  $X_2$ .

Если для анализируемых выборок  $Y_1$  и  $Y_2$  при помощи критерия согласия хи-квадрат (Пирсона) обосновано, что они распределены по нормальному закону, то для количественной оценки типичности результативного признака можно использовать выборочные средние арифметические значения  $\bar{Y}_1$  и  $\bar{Y}_2$ , а для оценки вариации – выборочные средние квадратичные отклонения  $S_1$  и  $S_2$  или выборочные дисперсии  $S_1^2$  и  $S_2^2$  [1]. Результат сравнения выборочных средних значений  $\bar{Y}_1$  и  $\bar{Y}_2$ , позволит обоснованно утверждать о влиянии факторного признака  $X$  на результативный признак  $Y$ . Формальная процедура сравнения выборочных средних значений  $\bar{Y}_1$  и  $\bar{Y}_2$  начинается с формулировки нулевой и альтернативной гипотез. Пусть, например, выборочное среднее значение  $\bar{Y}_1$ , вычисленное по выборке  $Y_1$ , оказалось больше, чем выборочное среднее значение  $\bar{Y}_2$ , вычисленное по выборке  $Y_2$ . Важно обосновать, является ли это различие действительно значимым, или это различие объясняется ошибкой выборки и не выходит за рамки статистической ошибки. Для этого формулируется статистическая гипотеза  $H_0$  (нулевая гипотеза), состоящая в том, что выборочные средние значения  $\bar{Y}_1$  и  $\bar{Y}_2$  значимо не различаются между собой, а наблюдаемые различия между их числовыми значениями являются результатом статистической ошибки, которая всегда имеет место при проведении выборочных исследований. Альтернативная односторонняя гипотеза  $H_1$  состоит в том, что выборочные средние значения значимо различаются

и значение  $\bar{Y}_1$  значимо больше, чем значение  $\bar{Y}_2$ , т.е. наблюдаемые различия между выборочными числовыми значениями выходят за рамки статистической ошибки. Проверка гипотезы  $H_0$  может быть произведена при помощи  $t$ -критерия. По своей сути он представляет собой математический алгоритм, в результате выполнения которого принимается решение о состоятельности гипотезы  $H_0$ . Следует отметить, что статистический критерий не используется для доказательства нулевой гипотезы. Он направлен на её опровержение. При этом может возникнуть ошибка первого рода. Ошибка первого рода – это ошибочное решение  $H_1/H_0$ , в результате которого проверяемая гипотеза  $H_0$  отклоняется и принимается альтернативная гипотеза  $H_1$  (числитель), когда на самом деле верна гипотеза  $H_0$  (знаменатель). Вероятность ошибки первого рода  $\alpha = P\left(\frac{H_1}{H_0}\right)$  также называется уровнем значимости. Одновременно уровень значимости однозначно определяет специфичность  $1 - \alpha = P\left(\frac{H_0}{H_0}\right)$ , которая количественно выражает вероятность правильного решения  $H_0/H_0$ , состоящего в том, что принимается гипотеза  $H_0$  (числитель) и она действительно верна (знаменатель). Для проверки нулевой гипотезы фиксируется вероятность ошибки первого рода и специфичность. Значение ошибки первого рода выбирается близким к нулю, при этом специфичность принимает значение близкое к единице. Обычно для проверки статистических гипотез в медико-биологических и фармакологических исследованиях выбирается уровень значимости  $\alpha = 0,05$  и специфичность  $1 - \alpha = 0,95$ .

Кроме вероятности ошибки первого рода  $\alpha$  и связанной с ней специфичности  $1 - \alpha$  необходимо задать приемлемую для проводимого статистического исследования фиксированную вероятность ошибки второго рода  $\beta$  и соответствующую ей мощность  $1 - \beta$  статистического критерия. На практике приемлемой обычно считается мощность статистического критерия  $1 - \beta$ , равная 0,8. Это соответствует вероятности ошибки второго рода  $\beta = 0,2$ . Это условие для вероятности ошибки второго рода является следствием соотношения «один к четырем» между выбранным уровнем значимости  $\alpha$  и вероятностью ошибки второго рода  $\beta$ . Для выбранного уровня значимости  $\alpha = 0,05$  фиксированная вероятность ошибки второго рода равна  $\beta = 4 \cdot \alpha = 4 \cdot 0,05 = 0,2$ . Тогда фиксированная мощность статистического критерия составит  $1 - \beta = 1 - 0,2 = 0,8$ .

Существует 2 варианта реализации  $t$ -критерия. Первый вариант используется в том случае, когда выборочные дисперсии  $S_1^2$  и  $S_2^2$  значимо не различаются. При помощи критерия Фишера [2] можно проверить статистическую гипотезу  $H_0$  об отсутствии значимых различий между выборочными дисперсиями  $S_1^2$  и  $S_2^2$  при альтернативной гипотезе  $H_1$ , которая состоит в том, что одна из выборочных дисперсий значимо больше другой. Сущность критерия Фишера и порядок вычислений достаточно подробно представлены в работе [2]. Статистика критерия  $t_1$  и требуемые для ее нахождения величины вычисляются по следующим формулам:

$$t_1 = \frac{|\bar{Y}_1 - \bar{Y}_2|}{S_x \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}, \quad \text{где} \quad S_x = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}}, \quad (1)$$

$$S_1^2 = \frac{\sum_{i=1}^{N_1} (y_{1i} - \bar{Y}_1)^2}{N_1 - 1}, \quad S_2^2 = \frac{\sum_{i=1}^{N_2} (y_{2i} - \bar{Y}_2)^2}{N_2 - 1}, \quad \bar{Y}_1 = \frac{\sum_{i=1}^{N_1} y_{1i}}{N_1}, \quad \bar{Y}_2 = \frac{\sum_{i=1}^{N_2} y_{2i}}{N_2}.$$

В приведенных формулах величины  $y_{1i}$  и  $y_{2i}$  представляют собой числовые значения выборочных множеств  $Y_1$  и  $Y_2$  соответственно. Статистика  $t_1$  в этом случае имеет распределение Стьюдента с числом степеней свободы  $k_1 = N_1 + N_2 - 2$ .

Второй вариант  $t$ -критерия используется в том случае, когда выборочные дисперсии  $S_1^2$  и  $S_2^2$  значимо различаются. Задача сравнения средних значений двух нормально распределённых выборок при значимо различающихся дисперсиях  $S_1^2$  и  $S_2^2$  известна как проблема Беренса–Фишера. Точного решения этой задачи до настоящего времени нет. На практике используются

различные приближенные вычисления. Особенность второго варианта  $t$ -критерия состоит в том, что статистика критерия  $t_2$  имеет распределение условно близкое к распределению Стьюдента с числом степеней свободы  $k_2$ . Для вычисления статистики критерия  $t_2$  и числа степеней свободы  $k_2$  используются следующие формулы:

$$t_2 = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{S_1}{N_1} + \frac{S_2}{N_2}}}, \quad k_2 = \frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{\left(\frac{S_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{S_2^2}{N_2}\right)^2}{N_2 - 1}} \quad (2).$$

В обоих вариантах  $t$ -критерия статистики критерия  $t_1$  и  $t_2$  количественно характеризуют различие между выборочными средними значениями  $\bar{Y}_1$  и  $\bar{Y}_2$ , при этом, чем больше эти статистики, тем менее вероятна нулевая гипотеза. Число степеней свободы является параметром распределения Стьюдента и заданное числовое значение этого параметра определяет функцию плотности вероятности  $f(t)$ , соответствующую этому распределению. В первом варианте вычислений статистика  $t_1$  имеет распределение Стьюдента с  $k_1$  степенями свободы. Во втором варианте  $t$ -критерия статистика  $t_2$  имеет приближенное распределение Стьюдента с числом степеней свободы  $k_2$ , поэтому последующие вычисления для этого варианта имеют в определенной степени приближенный характер. Дальнейшие действия, направленные на проверку нулевой гипотезы, вычисление ошибки второго рода, мощности критерия, аналогичны для обоих вариантов  $t$ -критерия. Поэтому расчетные значения статистик  $t$ -критерия  $t_1$  и  $t_2$  будем обозначать через  $t_p$ . В первом варианте вычислений  $t$ -критерия под значением  $t_p$  будем понимать значение  $t_1$ , т.е.  $t_p = t_1$ , а во втором варианте под значением  $t_p$  будем понимать  $t_2$ , т.е.  $t_p = t_2$ .

Величина  $p$ , характеризующая вероятность нулевой гипотезы, вычисляется из уравнения

$$p = \int_{t_p}^{\infty} f(t) dt, \quad \text{где } f(t) \text{ – плотность вероятности распределения Стьюдента с числом степеней}$$

свободы  $k_1$  или  $k_2$ . Число степеней свободы выбирается в зависимости от варианта вычислений  $t$ -критерия. Для автоматизации вычислений интеграла и нахождения значения  $p$  может быть использована, например, встроенная статистическая функция СТЬЮДРАСП табличного процессора Excel 2007 или аналогичные статистические функции этого табличного процессора последующих версий. Также для проверки нулевой гипотезы при односторонней альтернативной гипотезе из уравнения  $\alpha = \int_{t_{кр}}^{\infty} f(t) dt$  может быть вычислена величина  $t_{кр}$ . Для автоматизации

вычислений данного уравнения и нахождения значения  $t_{кр}$  может быть использована статистическая функция СТЬЮДРАСПОБР табличного процессора Excel 2007 или аналогичные статистические функции последующих версий этой программы. Вычисленные значения  $p$  и  $t_{кр}$  необходимы для проверки (опровержения) нулевой гипотезы. При принятии решения все вероятности, меньшие или равные  $\alpha$  считаются малыми, а разность между выборочными средними значениями  $\bar{Y}_1 - \bar{Y}_2$  значимой, если  $t_p$  больше или равна  $t_{кр}$ . Обычно в большинстве научных публикаций принятие решения в отношении гипотезы  $H_0$  при зафиксированной ошибке первого рода  $\alpha$  принимается в соответствии со следующим правилом. Если  $t_p \geq t_{кр}$  и, соответственно,  $p \leq \alpha$ , то есть основание отклонить нулевую гипотезу  $H_0$  и принять альтернативную  $H_1$ . Если  $t_p < t_{кр}$  и, следовательно,  $p > \alpha$ , то нет оснований отклонить нулевую гипотезу  $H_0$ . Следует отметить, что условия  $t_p \geq t_{кр}$  и  $p \leq \alpha$  формально взаимосвязаны и выполняются одновременно, также как и условия  $t_p < t_{кр}$  и  $p > \alpha$ .

Решение в отношении гипотезы  $H_0$ , принимаемое на основании приведенных выше правил, не учитывает вероятность ошибки второго рода и мощность критерия. Выводы в отношении гипотезы  $H_0$  можно считать состоятельными, когда правило, на основании которого принимается решение, учитывает как ошибку первого рода  $\alpha$ , так и мощность (чувствительность) критерия  $1 - \beta$ .

Для вычисления мощности используется функция плотности вероятности нецентрального распределения Стьюдента. Это достаточно сложная математическая функция и ее интегрирование представляет существенные трудности для исследователей, не имеющих специальной математической подготовки. Для автоматизации вычислений, связанных с этой функцией, необходимо использовать специальные программы, например систему компьютерной математики Maple. Например, популярный и часто используемый для статистической обработки экспериментальных данных табличный процессор Excel, даже последних версий, не содержит встроенных функций, автоматизирующих интегрирование нецентрального распределения Стьюдента.

Расчетное значение ошибки второго рода  $\beta_p$  может быть вычислено по формуле  $\beta_p = 1 - \int_{t_{кр}}^{\infty} s(t) dt$ ,

где  $s(t)$  – функция плотности вероятности нецентрального распределения Стьюдента. Для первого варианта  $t$ -критерия нецентральное распределение Стьюдента  $s(t)$  задается числом степеней свободы  $k_1$  и параметром нецентральности  $\lambda = t_1$ , а для второго варианта – числом степеней свободы  $k_2$  и параметром нецентральности  $\lambda = t_2$ . Соответственно, мощность критерия вычисляется по формуле  $1 - \beta_p = \int_{t_{кр}}^{\infty} s(t) dt$ . Для автоматизации вычислений можно использовать

электронный ресурс Keisan online calculator, который находится в открытом доступе по адресу <https://keisan.casio.com>.

С учетом вычисленных значений ошибки второго рода  $\beta_p$  и мощности критерия  $1 - \beta_p$  могут быть сформулированы следующие выводы в отношении проверяемой нулевой гипотезы  $H_0$ . В случае, когда при выполнении условия  $p \leq \alpha$  статистическая гипотеза  $H_0$  отклоняется, то принимаемая на этом основании статистическая гипотеза  $H_1$  является состоятельной и отражает реальные различия между средними значениями  $\bar{Y}_1$  и  $\bar{Y}_2$ , если ошибка второго рода  $\beta_p$  не превосходит зафиксированного в начале исследования значения  $\beta$ , а вычисленная мощность критерия  $1 - \beta_p$  не меньше фиксированной мощности  $1 - \beta$ . При исследовании малых выборок ситуация, когда  $p \leq \alpha$ , но  $\beta_p > \beta$  и  $(1 - \beta_p) < (1 - \beta)$ , указывает на большую ошибку второго рода и на недостаточную мощность критерия, которая может быть увеличена за счет увеличения объема выборочных совокупностей. Очевидно, что статистические различия можно считать значимыми и особенно надежными, если условия  $p \leq \alpha$  и  $(1 - \beta_p) \geq (1 - \beta)$  выполняются для выборок, объем которых меньше 30 значений. Кроме этого, отклонение гипотезы  $H_0$  при  $p \leq \alpha$ , высокой мощности  $(1 - \beta_p) \geq (1 - \beta)$  и малой ошибке второго рода  $\beta_p < \beta$  имеет большое значение, поскольку позволяет обосновать достаточность объема малой выборки. Это важно при проведении фармакологических исследований, когда малый объем исследуемых выборок обычно объективно обоснован. Следовательно, при анализе малых выборок в случае отклонения нулевой гипотезы ( $p \leq \alpha$ ) этот вывод можно считать состоятельным, если вычисленная ошибка второго рода  $\beta_p$  меньше зафиксированного значения  $\beta$ , а вычисленная мощность  $t$ -критерия  $1 - \beta_p$  больше зафиксированной мощности  $1 - \beta$ .

## Результаты исследования и их обсуждение

В результате проведенного теоретического исследования был разработан способ оценки мощности, который совместно с алгоритмом  $t$ -критерия позволяет уточнить выводы в отношении



проверяемой статистической гипотезы об отсутствии значимых различий между средними значениями нормально распределенных выборок. Разработанный способ может быть использован для двух вариантов  $t$ -критерия, один из которых используется для сравнения средних значений при значимо не различающихся дисперсиях, а другой – при значимо различающихся дисперсиях. Для автоматизации вычисления мощности и ошибки второго рода предложено использование доступного интернет-ресурса Keisan online calculator, который находится в открытом доступе по адресу <https://keisan.casio.com>. Разработанный способ содержит следующие этапы вычислений.

1. Для анализируемых выборочных совокупностей  $Y_1$  и  $Y_2$ , имеющих объемы соответственно  $N_1$  и  $N_2$ , по следующим формулам вычисляются выборочные средние значения  $\bar{Y}_1$  и  $\bar{Y}_2$ , а также выборочные значения несмещенных дисперсий  $S_1^2$  и  $S_2^2$ :

$$\bar{Y}_1 = \frac{\sum_{i=1}^{N_1} y_{1i}}{N_1}, \quad \bar{Y}_2 = \frac{\sum_{i=1}^{N_2} y_{2i}}{N_2}, \quad S_1^2 = \frac{\sum_{i=1}^{N_1} (y_{1i} - \bar{Y}_1)^2}{N_1 - 1}, \quad S_2^2 = \frac{\sum_{i=1}^{N_2} (y_{2i} - \bar{Y}_2)^2}{N_2 - 1}. \quad (3)$$

2. При помощи критерия согласия  $\chi^2$  (хи-квадрат) производится проверка статистической гипотезы о соответствии эмпирических распределений анализируемых выборок теоретическому нормальному распределению. Для анализа малых выборок целесообразно использовать равночастотную группировку эмпирических данных [1]. В качестве параметров нормального распределения используются выборочные параметры  $\bar{Y}_1$ ,  $S_1^2$  и  $\bar{Y}_2$ ,  $S_2^2$ . В случае, если для выборок  $Y_1$  и  $Y_2$  обоснованно принимается гипотеза о нормальности их распределения, то для проверки гипотезы об отсутствии значимых различий между средними значениями  $\bar{Y}_1$  и  $\bar{Y}_2$  этих выборок можно использовать  $t$ -критерий Стьюдента.

3. При помощи критерия Фишера [2] проверяется статистическая гипотеза об отсутствии значимых различий между выборочными дисперсиями  $S_1^2$  и  $S_2^2$ . Если выборочные дисперсии значимо не различаются, то для сравнения выборочных средних значений  $\bar{Y}_1$  и  $\bar{Y}_2$  применяется первый вариант вычислений  $t$ -критерия, а в случае значимых различий между ними – второй.

4. Для проверки статистической гипотезы об отсутствии значимых различий между средними значениями  $\bar{Y}_1$  и  $\bar{Y}_2$  фиксируется уровень значимости (ошибка первого рода)  $\alpha$  и специфичность критерия  $1 - \alpha$ , а также допустимая ошибка второго рода  $\beta$  и мощность критерия  $1 - \beta$ .

5. Выполняются вычисления по первому или второму варианту  $t$ -критерия. В результате определяются значения статистик  $t_1$  или  $t_2$ , а также соответствующие им степени свободы  $k_1$  или  $k_2$ .

6. В результате решения следующих уравнений производится вычисление вероятности  $p$  нулевой гипотезы  $H_0$  и критического значения статистики  $t_{кр}$ :

$$p = \int_{t_p}^{\infty} f(t) dt, \quad \alpha = \int_{t_{кр}}^{\infty} f(t) dt,$$

где  $f(t)$  – плотность распределения вероятности (распределение Стьюдента) с числом степеней свободы  $k_1$  или  $k_2$  в зависимости от выбранного варианта вычислений  $t$ -критерия. Также в первом варианте вычислений  $t$ -критерия  $t_p = t_1$ , а во втором варианте  $t_p = t_2$ . Для автоматизации вычислений могут быть использованы, например, встроенные статистические функции табличного процессора Excel.

7. Расчетное значение ошибки второго рода  $\beta_p$  и мощность критерия  $1 - \beta_p$  вычисляются по формулам:

$$\beta_p = 1 - \int_{t_{кр}}^{\infty} s(t) dt, \quad 1 - \beta_p = \int_{t_{кр}}^{\infty} s(t) dt,$$

где  $s(t)$  – функция плотности вероятности для нецентрального распределения Стьюдента. Для первого варианта  $t$ -критерия нецентрального распределения Стьюдента  $s(t)$  задается числом степеней свободы  $k_1$  и параметром нецентральности  $\lambda = t_1$ , а для второго варианта – числом

степеней свободы  $k_2$  и параметром нецентральности  $\lambda = t_2$ . Для автоматизации вычисления можно использовать электронный ресурс Keisan online calculator, который находится в открытом доступе по адресу <https://keisan.casio.com>.

8. Решение об отклонении гипотезы  $H_0$  принимается при выполнении следующих условий:  $t_p \geq t_{кр}$ ,  $p \leq \alpha$  и  $\beta_p < \beta$  и  $(1 - \beta_p) \geq (1 - \beta)$ . Вероятность нулевой гипотезы  $p$  не должна быть больше зафиксированной ошибки первого рода  $\alpha$ , а мощность критерия  $1 - \beta_p$  не должна быть меньше зафиксированной мощности  $1 - \beta$ . Если указанные условия выполняются, то вывод об отклонении гипотезы  $H_0$  и принятие альтернативной гипотезы  $H_1$  можно считать состоятельным даже при малых объемах выборочных совокупностей. Если условия  $t_p \geq t_{кр}$ ,  $p \leq \alpha$  выполняются, но мощность критерия недостаточна  $(1 - \beta_p) < (1 - \beta)$ , то это указывает на необходимость проведения дополнительных исследований с целью увеличения объема выборок и повышения мощности критерия. Постепенное увеличение объема выборок и проверка мощности  $t$ -критерия позволит оптимизировать проведение экспериментов и обеспечить экономию времени и средств.

*Пример практического применения  $t$ -критерия с оценкой его мощности.* Рассмотрим вычисление мощности  $t$ -критерия на практическом примере. В приведенной ниже последовательности действий непосредственные вычисления, связанные с нахождением выборочных средних, дисперсий, проверкой выборок на нормальность, проверкой равенства дисперсий, не приводятся, но показаны результаты этих вычислений. Это сделано с целью обратить основное внимание непосредственно на  $t$ -критерий, оценку его мощности и формулировку выводов, но вместе с тем отразить и общий порядок вычислений.

Пусть в результате проведенных экспериментов получены две выборочные совокупности  $Y_1$  и  $Y_2$ . Каждая из выборок содержит по десять значений, т.е.  $N_1 = N_2 = 10$ . Множества числовых значений выборок  $Y_1$  и  $Y_2$  количественно выражают состояние единиц наблюдения, вызванные факторными воздействиями  $X_1$  и  $X_2$ .

1. Выборочные средние значения и дисперсии для анализируемых выборок, вычисленные по формулам (3), соответственно равны:  $\bar{Y}_1 = 151,9$ ;  $S_1^2 = 12,25$ ;  $\bar{Y}_2 = 155,6$ ;  $S_2^2 = 5,29$ .
2. В силу малого объема анализируемых выборок непосредственно проверить статистическую гипотезу о нормальном распределении выборок не представляется возможным. Допустим, что имеются теоретические основания и опубликованные экспериментальные результаты, позволяющие полагать, что в данных экспериментах значения результативного признака в анализируемых выборках распределены по нормальному закону.
3. В результате проверки статистической гипотезы об отсутствии значимых различий между выборочными дисперсиями  $S_1^2$  и  $S_2^2$  при помощи критерия Фишера на уровне значимости  $\alpha = 0,05$  установлено, что выборочные дисперсии значимо не различаются.

$$F_p = \frac{S_1^2}{S_2^2} = \frac{12,25}{5,29} = 2,32; \quad p = 0,11 > 0,05.$$

Следовательно, можно использовать первый вариант  $t$ -критерия.

4. Для проверки гипотезы об отсутствии значимых различий между выборочными средними значениями зафиксируем ошибку первого рода  $\alpha = 0,05$  и ошибку второго рода  $\beta = 4 \cdot \alpha = 4 \cdot 0,05 = 0,2$ . Соответственно специфичность  $t$ -критерия равна  $1 - \alpha = 0,95$ , а мощность фиксируется на уровне  $1 - \beta = 1 - 0,2 = 0,8$ .
5. Вычислим статистику  $t$ -критерия  $t_p = t_1$  по формулам (1).

$$S_x = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}} = \sqrt{\frac{9 \cdot 12,25 + 9 \cdot 5,29}{18}} = 2,96; \quad t_p = \frac{|\bar{Y}_1 - \bar{Y}_2|}{S_x \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{|151,9 - 155,6|}{2,96 \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}} = 2,8.$$

6. Вычислим вероятность  $p$  нулевой гипотезы  $H_0$  и критическое значение статистики  $t_{кр}$ . Для автоматизации вычислений воспользуемся табличным процессором Microsoft Excel. Статистика  $t$ -критерия  $t_p$  имеет теоретическое распределение Стьюдента  $f(t)$ . Число степеней свободы, полностью определяющее эту функцию, равно  $k_1 = N_1 + N_2 - 2 = 10 + 10 - 2 = 18$ . Для автоматизации вычислений интеграла  $p = \int_{t_p}^{\infty} f(t) dt$  воспользуемся статистической функцией СТЬЮДРАСП табличного процессора Excel 2007. Результат вычислений:  $p = \text{СТЬЮДРАСП}(2,8; 18; 1) = 0,006$ . Для решения уравнения  $\alpha = \int_{t_{кр}}^{\infty} f(t) dt$  при условии, что  $\alpha = 0,05$ , воспользуемся функцией СТЬЮДРАСПОБР. Данная функция вычисляет критическое значение  $t_{кр}$  для двух областей  $t < -t_{кр}$  и  $t > t_{кр}$  распределения Стьюдента. Поэтому при задании функции в поле «вероятность» необходимо указывать значение  $2 \cdot \alpha = 0,1$ . Результат вычислений:  $t_{кр} = \text{СТЬЮДРАСПОБР}(0,1; 18) = 1,73$ .

7. Вычислим расчетные значения мощности  $t$ -критерия  $1 - \beta_p = \int_{t_{кр}}^{\infty} s(t) dt$  и ошибки второго рода  $\beta_p$ .

В данном интеграле функция  $s(t)$  – нецентральное распределение Стьюдента с числом степеней свободы  $k_1 = 18$  и параметром нецентральности  $\lambda = t_p = 2,8$ . Для автоматизации вычисления воспользуемся электронным калькулятором Keisan online calculator, который находится в открытом доступе по адресу <https://keisan.casio.com>. (рис.)

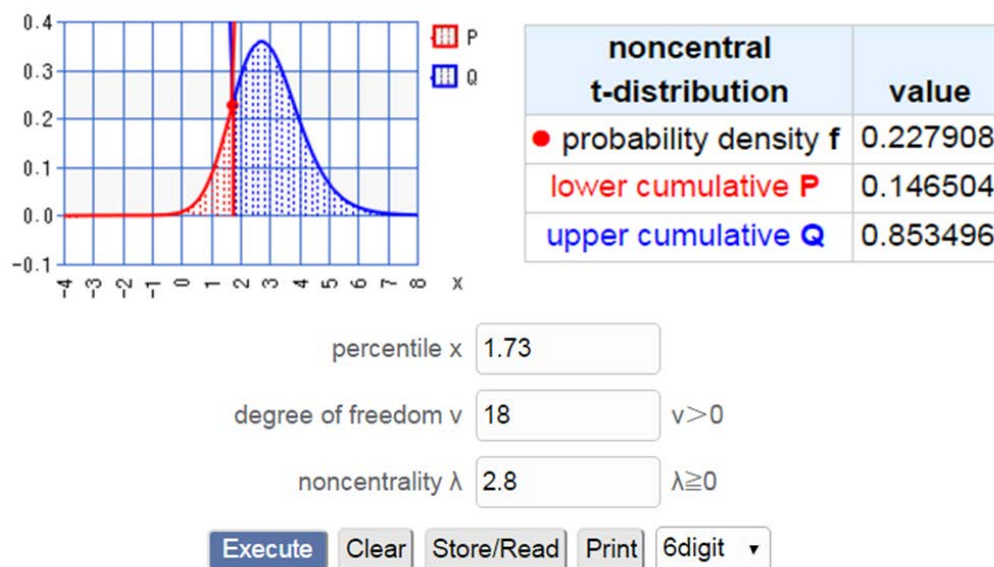


Рис. Рабочий лист калькулятора Keisan online calculator

На рабочем листе калькулятора в поле «percentile» вводится критическое значение статистики  $t_{кр} = 1,73$ . В поле «degree of freedom» указывается число степеней свободы  $k_1 = 18$ . В поле «noncentrality» заносится значение параметра нецентральности  $\lambda = t_p = 2,8$ . После нажатия кнопки «Execute» калькулятор вычисляет расчетную вероятность ошибки второго рода  $\beta_p = 0,15$  (lower cumulative P) и расчетную мощность  $t$ -критерия  $1 - \beta_p = 0,85$  (upper cumulative Q).

8. Поскольку в результате вычислений установлено, что  $(t_p = 2,8) > (t_{кр} = 1,73)$  и  $(p = 0,006) < (\alpha = 0,05)$

, то это значит, что есть основания отклонить гипотезу  $H_0$  и принять гипотезу  $H_1$ . Среднее значение  $\bar{Y}_2 = 155,6$  выборки  $Y_2$  значимо больше среднего значения  $\bar{Y}_1 = 151,9$  выборки  $Y_1$ . При этом расчетная вероятность ошибки второго рода  $\beta_p = 0,15$  меньше зафиксированного в начале исследования значения  $\beta = 0,2$ . Расчетная мощность  $t$ -критерия  $1 - \beta_p = 0,85$  больше зафиксированной мощности  $1 - \beta = 0,8$ . Следовательно, вывод о значимом различии выборочных средних значений можно считать состоятельным, несмотря на малый объем исследуемых выборочных совокупностей.

## Заключение

Разработанный способ вычисления мощности является логическим завершением  $t$ -критерия и позволяет рассматривать его как законченную процедуру принятия статистического решения в отношении проверяемой гипотезы. Предложенный порядок вычисления и вариант их автоматизации позволяет достаточно просто и без больших временных затрат провести статистический анализ и обосновать состоятельность статистических выводов, поскольку в результате проведенных вычислений становятся известными все статистические ошибки (первого и второго рода). Решение в отношении проверяемой гипотезы принимается на основании сравнения вычисленных вероятностей и статистических ошибок с заданными в начале исследования значениями. Предложенный способ вычислений может представлять интерес для исследователей, не имеющих специальной математической подготовки, но использующих в своей работе методы математической статистики.

## Литература (references)

1. Лямец Л.Л., Евсеев А.В. Методика проверки гипотезы о нормальном распределении малой выборки в фармакологических исследованиях // Вестник Смоленской государственной медицинской академии. – 2019. – №1, Т.18. – С. 55-66. [Ljamec L.L., Evseev A.V. *Vestnik Smolenskoj gosudarstvennoj medicinskoj akademii*. Bulletin of the Smolensk state medical Academy. – 2019. – N1, V.18. – P. 55-66. (in Russian)]
2. Медик В.А., Токмачев М.С., Фишман Б.Б. Статистика в медицине и биологии: Руководство. В 2-х томах / Под редакцией Ю.М. Комарова. Т.1. Теоретическая статистика. – М.: Медицина, 2000. – 412 с. [Medik V.A., Tokmachev M.S., Fishman B.B. *Statistika v medicine i biologii: Rukovodstvo. T.1. Teoreticheskaja statistika*. Statistics in medicine and biology: a Guide. V.1 Theoretical statistics. – Moscow: Medicine, 2000. – 412 p. (in Russian)]
3. Гланц С. Медико-биологическая статистика. Пер. с англ. – М.: Практика, 1998. – 459 с. [Glanc S. *Mediko-biologicheskaja statistika*. Medico-biological statistics. – Moscow: Praktika, 1998. – 459 p. (in Russian)]
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2003. – 350 с. [Sidorenko E.V. *Metody matematicheskoy obrabotki v psihologii*. Methods of mathematical processing in psychology. – Saint-Petersburg: LLC "Speech", 2003. – 350 p. (in Russian)]

## Информация об авторах

Лямец Леонид Леонидович – кандидат технических наук, доцент, заведующий кафедрой физики, математики и медицинской информатики ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: LLL190965@yandex.ru

Евсеев Андрей Викторович – доктор медицинских наук, профессор, заведующий кафедрой нормальной физиологии ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: hypoxia@yandex.ru

Козлова Евгения Карповна – старший преподаватель кафедры физики, математики и медицинской информатики ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: fmmi@smolgm.ru

Колпакова Марина Анатольевна – старший преподаватель кафедры физики, математики и медицинской информатики ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России. E-mail: fmmi@smolgm.ru