

УДК 614.2

<https://doi.org/10.22328/2413-5747-2023-9-2-111-125>

РАСЧЕТ НЕОБХОДИМОГО ОБЪЕМА ВЫБОРКИ С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ G*POWER

*Е.А. Кригер, С.Н. Драчев, Н.А. Митькин, В.А. Постоев, А.М. Гржибовский**
Северный государственный медицинский университет, г. Архангельск, Россия

Расчет адекватного объема выборки является важнейшим шагом в получении достоверных и обобщаемых результатов в сфере биомедицинских исследований. Это позволяет найти компромисс между временными, ресурсными и этическими аспектами планирования и проведения исследований. Статистическое программное обеспечение позволяет быстро и удобно рассчитать размер выборки, но выбор метода расчета в значительной степени зависит от характеристик данных и исследовательского вопроса.

В статье представлен практический и доступный метод расчета необходимого объема выборки для проведения исследований, ставящих целью сравнение средних величин в нескольких группах при помощи бесплатного программного обеспечения G*Power.

Описаны варианты расчета выборки для применения критерия Стьюдента для двух независимых выборок и однофакторного дисперсионного анализа, а также непараметрических критериев Вилкоксона и Манна-Уитни. Статья шаг за шагом проводит читателей через расчеты объема выборки для сравнения одной группы с заданным значением, двух независимых групп, а также трех и более независимых групп. Последовательное изложение материала и формат примеров в статье облегчают знакомство читателей с новой и доступной программой. Настоящая работа может быть полезна начинающим исследователям, находящимся на этапе планирования исследований, как практическое руководство по расчету необходимого объема выборки для параметрических и непараметрических методов анализа с целью проверки статистических гипотез о равенстве центральных тенденций в независимых совокупностях.

КЛЮЧЕВЫЕ СЛОВА: морская медицина, размер выборки, параметрические критерии, непараметрические критерии, G*Power

*Для корреспонденции: Гржибовский Андрей Мечиславович, e-mail: A.Grjibovski@yandex.ru

*For correspondence: *Andrej M. Grjibovski*, e-mail: A.Grjibovski@yandex.ru

Для цитирования: Кригер Е.А., Драчев С.Н., Митькин Н.А., Постоев В.А., Гржибовский А.М. Расчет необходимого объема выборки с использованием программы G*Power // *Морская медицина*. 2023. Т. 9, № 2. С. 111-125, doi: <https://doi.org/10.22328/2413-5747-2023-9-2-111-125> EDN: <https://elibrary.ru/ZDCEUG>

For citation: Krieger E.A., Drachev S.N., Mitkin N.A., Postoev V.A., Grjibovski A.M. Sample size calculation using G*Power software // *Marine Medicine*. 2023. Vol. 9, № 2. P. 111-125, doi: <https://doi.org/10.22328/2413-5747-2023-9-2-111-125> EDN: <https://elibrary.ru/ZDCEUG>

SAMPLE SIZE CALCULATION USING G*POWER SOFTWARE

*Ekaterina A. Krieger, Sergei N. Drachev, Nikita A. Mitkin, Vitaly A. Postoev, Andrej M. Grjibovski**
Northern State Medical University, Arkhangelsk, Russia

Sample size calculation is a critical step in planning a research project to obtain valid and generalizable results in biomedical sciences. It enables researchers to reach a balance between time, resources, and ethical considerations when planning and conducting research. Statistical software can assist in calculating sample size, but calculation method depends on the data characteristics and the research question.

This article presents a practical tool for calculating the required sample size for studies that aim to compare measures of central tendency for several research questions using the G*Power software. Sample size calculation for Student's unpaired t-test, one-way analysis of variance (ANOVA), as well as their non-parametric analogs – Wilcoxon and Mann-Whitney tests are presented. The article walks readers step by step through sample size calculations for the abovementioned situations using practical examples. Therefore, this paper has a potential to become a valuable resource for junior researchers who are at the planning

© Авторы, 2023. Издатель ООО Балтийский медицинский образовательный центр. Данная статья распространяется на условиях «открытого доступа», в соответствии с лицензией ССВУ-NC-SA 4.0 («Attribution-NonCommercialShare-Alike» / «Атрибуция-Некоммерчески-Сохранение Условий» 4.0), которая разрешает неограниченное некоммерческое использование, распространение и воспроизведение на любом носителе при условии указания автора и источника. Чтобы ознакомиться с полными условиями данной лицензии на русском языке, посетите сайт: <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.ru>

phase of their research. It serves as a practical guide for calculating the required sample size for parametric and nonparametric methods of analysis to test statistical hypotheses about the equality of measures of central tendency in unpaired groups.

KEYWORDS: marine medicine, sample size, parametric criteria, non-parametric criteria, G*Power.

Введение. При биомедицинских исследованиях, оценивая параметры генеральной совокупности (например, среднее или медианное значение количественного показателя; долю; относительный риск; отношение шансов и др.), часто полагаются на результаты, полученные на выборочной совокупности (выборке), которая сформирована из этой генеральной совокупности. Ключевым вопросом на этапе планирования таких исследований является обеспечение качественной и количественной репрезентативности данной выборки. Последняя определяется минимальным количеством участников исследования (объемом выборки), необходимым для ответа на поставленный исследовательский вопрос с определенным уровнем точности и приемлемой вероятностью ошибок 1-го и 2-го рода [1]. Действительно, недостаточный объем выборки не позволит исследователю обнаружить различия между изучаемыми группами, даже если в действительности эти различия существуют. С другой стороны, слишком большая выборочная совокупность приведет к неэффективному расходованию финансовых и временных ресурсов на ее формирование; большее число участников исследования будет подвергнуто неоправданному риску, что особенно важно в экспериментальных исследованиях, и является неоправданным с этической точки зрения [1]. Кроме того, статистически значимые различия между исследуемыми группами, полученные на больших выборках, могут не иметь клинической значимости, что также необходимо учитывать при анализе результатов и формулировании выводов.

В исследовательской практике для сравнения средних значений используются параметрические методы, такие как *t*-критерий Стьюдента и дисперсионный анализ [2, 3]. Однако данные в биомедицинских исследованиях могут иметь распределение, отличное от нормального, что делает параметрические методы ненадежными, особенно при малых объемах выборки. В этом случае исследователи могут рассмотреть альтернативные методы, такие как преобразование данных или непараметрические критерии [4].

Целью данной работы является пошаговое описание расчета объема выборки для сравнения центральных тенденций (средних величин) с использованием как параметрических, так и непараметрических статистических критериев с помощью программного обеспечения G*Power в помощь исследователям, планирующим свои научные проекты. Расчеты объема выборки будут представлены для непарного критерия Стьюдента и однофакторного дисперсионного анализа трех и более независимых групп; рангового критерия Вилкоксона для одной выборки и *U*-критерия Манна–Уитни для двух независимых выборок.

Существуют разные подходы к расчету объема выборки, которые могут опираться либо на приемлемую точность результатов исследования, либо на статистическую мощность исследования. Второй подход встречается более часто в аналитических исследованиях, и большинство современных средств автоматизированного определения объема выборки используют именно его. Хотя данный расчет может быть реализован вручную, с использованием формул, которые приведены в пособиях по статистическому анализу, более удобным и быстрым способом расчета является применение пакетов прикладных статистических программ, таких как Stata (StataCorp, College Station, TX, США), STATISTICA (StatSoft, Россия), или специальных программ, таких как PASS (NCSS, Kaysville, Utah, США). Однако эти программы не являются бесплатными, что ограничивает их использование. В данной работе мы представляем возможности свободно распространяемого программного обеспечения G*Power для расчета объема выборки. Данный программный продукт разработан в университете Дюссельдорфа (Германия) и доступен к бесплатному скачиванию с официального сайта (<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>) [5, 6]. Программа может быть установлена на компьютеры, использующие операционные системы Windows и Mac OS. Для подготовки данной статьи была

использована программа G*Power 3.1.9.7 for Windows.

Расчет объема выборки для применения критерия Стьюдента для двух независимых групп

После установки программы G*Power для ее запуска необходимо нажать на ярлык **βa**. Откроется главное диалоговое окно, которое показано на рис. 1.

Для расчета объема выборки в графе Test family / Семейство критериев выбирается нужная категория статистического критерия из выпадающего списка, включающего Exact (точный)-тест, F-тест, t-тест, χ^2 -тест, z-тест. В нашем случае для сравнения двух независимых выборок, в которых распределение признака подчиняется закону нормального распределения, t-критерий Стьюдента для двух непарных выборок. Это параметрический тест, следовательно, в графе Statistical test / Статистический тест из выпадающего списка нужно выбрать Means: Difference between two independent means (two groups) / Средние: разница между двумя независимыми средними (две группы). Объем выборки необходимо рассчитывать при планировании исследования, то есть до начала процедуры сбора данных. В этом случае в графе Type of power analysis / Тип анализа мощности следует выбрать A priori: Compute required sample size – given α , power, and effect size / Априори: вычислить необходимый размер выборки — учитывая величину α -ошибки, мощность и величину эффекта. Далее в разделе Input parameters / Входные параметры необходимо задать 5 параметров для расчета: tail(s), α err prob, power (1- β err prob), effect size d, allocation ratio N2/N1. Остановимся на данных параметрах более подробно.

В программе G*Power используется принцип тестирования нулевой (H_0) гипотезы Неймана–Пирсона [7], поэтому перед проведением расчетов необходимо сформулировать H_0 и альтернативную (H_1) гипотезу. В зависимости от формулировки H_0 и H_1 в графе Tail(s) следует выбрать один из вариантов: one – односторонний тест или two – двусторонний тест. Пример формулировки H_0 и H_1 для сравнения средних значений какого-либо количественного показателя для двустороннего теста (two-tailed test) может быть записан следующим образом:

$H_0: \mu_1 = \mu_2$, среднее значение в одной популяции равно среднему значению в другой популяции;

$H_1: \mu_1 \neq \mu_2$, популяционные средние значения не равны, среднее значение в популяции 2 больше или меньше среднего значения в популяции 1.

Для одностороннего теста (one-tailed test) H_0 и H_1 гипотезы могут быть представлены как:

$H_0: \mu_1 \leq \mu_2$, среднее значение в популяции 1 не превышает среднее значение в популяции 2;

$H_1: \mu_1 > \mu_2$, среднее значение в популяции 1 выше среднего значения в популяции 2.

Применение одностороннего теста может быть использовано, например, при оценке эффективности лечения. Следует обратить внимание, что для проведения одностороннего теста потребуется меньший объем выборки в сравнении с двусторонним тестом, но использовать его следует только в том случае, если мы точно знаем направленность изучаемой взаимосвязи. Во всех остальных случаях применение двустороннего теста является вариантом выбора. Применение принципа Неймана–Пирсона для тестирования H_0 также реализовано в программе G*Power графически (рис. 2). В раз-

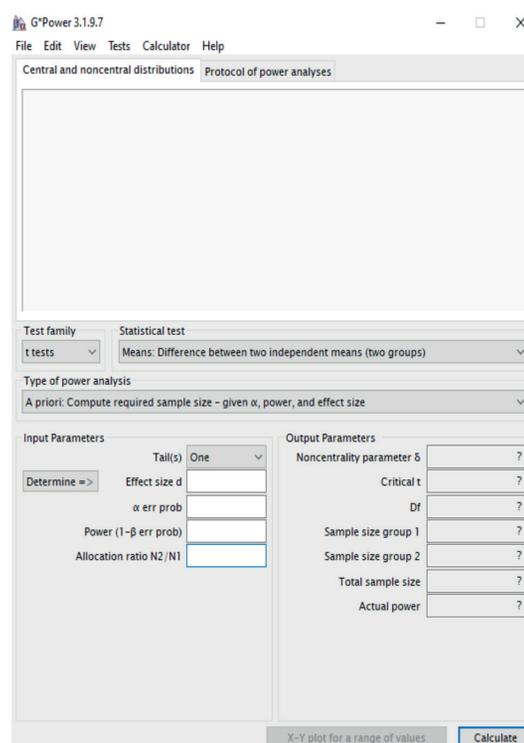


Рис. 1. Главное диалоговое окно программы G*Power

Fig. 1. G*Power main dialogue box

деле Central and noncentral distributions/ Центральное и нецентральное распределение (после введения всех необходимых параметров) будут представлены две кривые, демонстрирующие распределение значений t -критерия Стьюдента при условии, что верна H_0 (сплошная кривая) или верна H_1 (пунктирная кривая). Впоследствии при проведении статистического анализа данных мы принимаем одно из возможных решений: принимаем H_0 / отклоняем H_1 или отклоняем H_0 / принимаем H_1 . Таким образом, при отклонении нулевой гипотезы альтернативная гипотеза принимается автоматически и наоборот. Наш выбор будет зависеть от заданного уровня α -/ β -ошибок и мощности исследования, которые также следует учесть при расчете необходимого объема выборки.

Ошибка 1-го рода (α -ошибка) сопряжена с получением ложноположительного результата, то есть с вероятностью найти различия там, где они реально отсутствуют. Ошибка 2-го рода (β -ошибка) приводит к ложноотрицательному выводу, когда исследователи не находят различий там, где эти различия действительно существуют. И первая, и вторая ситуации являются нежелательными с точки зрения корректности будущих выводов. Традиционно в биомедицинских исследованиях величина α -ошибки зада-

ется равной 0,05 (или 5 %), а величина β -ошибки – 0,20 (или 20 %), но в зависимости от цели исследования размер этих ошибок может быть изменен. При этом важно учитывать, что при уменьшении уровня α -ошибки, увеличивается величина β -ошибки и наоборот, что наглядно представлено на рис. 2. Мощность исследования рассчитывается по формуле: $1-\beta$, то есть определяется вероятностью не совершить ошибку 2-го рода. Другими словами, статистическая мощность – вероятность найти различия, если они в действительности существуют [1]. Исходя из рекомендуемого уровня β -ошибки 0,20 (или 20 %), мощность исследования признается достаточной при значении 0,80 (80 %) и более. С учетом сказанного выше, для расчета объема выборки задается значение α -ошибки (α err prob) и необходимая мощность исследования Power ($1-\beta$ err prob).

Объем выборки зависит от величины различий между группами (величина эффекта), которую необходимо определить с учетом цели исследования и предположения о потенциальной клинической/эпидемиологической значимости этого эффекта. Важно понимать, что, с одной стороны, увеличение потенциального эффекта, который может быть зафиксирован в исследовании, будет снижать необходимый объем выборки, с другой стороны, арифметически небольшой эффект не обязательно будет иметь клиническую значимость и, как следствие, практическое применение. Значение величины эффекта может быть задано априори в поле Effect size d , где d – величина эффекта Коэна. Значение величины эффекта равно 0,2 считается малым эффектом, 0,5 – средним эффектом, 0,8 – большим эффектом. Кроме того, величину эффекта можно определить расчетным способом, нажав на Determine/Определить и указав среднее значение и стандартное отклонение (SD) для двух групп, опираясь на данные публикаций или небольшого пилотного исследования. Последним параметром, который необходимо ввести в поле Input parameters, является Allocation ratio N_2/N_1 / Коэффициент распределения N_2/N_1 , отражающий соотношение участников исследования в двух сравниваемых группах.

Рассмотрим, как проводится расчет необходимого объема выборки для сравнения количественных показателей, подчиняющихся закону нормального распределения, в двух независи-

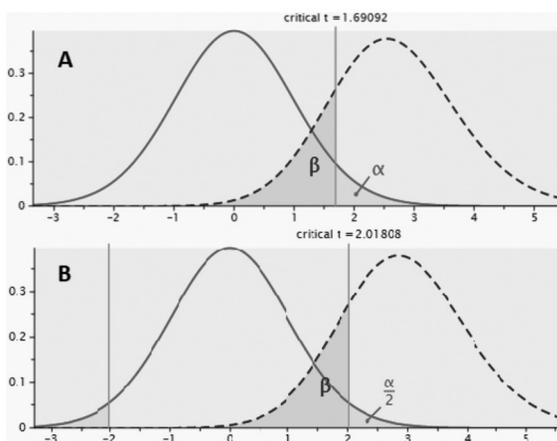


Рис. 2. Распределение значений t -критерия Стьюдента при условии, что верна нулевая гипотеза (сплошная кривая) или верна альтернативная гипотеза (пунктирная кривая) для А – одностороннего теста, В – двустороннего теста.

Fig. 2. T-values distribution under assumption of true null hypothesis (continuous curve) or true alternative hypothesis (dotted curve) for A – one-tailed test, B – two-tailed test

мых выборках на конкретном примере. Допустим, нас интересуют различия в среднем количестве калорий, потребляемых в сутки лицами разного пола. Заполняем соответствующие графы, как показано на рис. 3. Расчет Объем выборки будем рассчитывать для двустороннего теста, при уровне α -ошибки 0,05, мощности 0,80 и соотношения участников исследования в двух группах как 1 : 1. Предположим, что мы планируем определить среднюю величину эффекта, поэтому в графу Effect size d необходимо ввести величину 0,5.

После нажатия кнопки Calculate/Рассчитать получаем результат (рис. 4), который появится в разделе Output parameters / Выходные параметры и на основании которого мы можем сделать вывод о том, что в данное исследование необходимо включить минимум 128 человек (Total sample size): 64 мужчины (Sample size group 1) и 64 женщины (Sample size group 2).

Допустим, что с использованием мобильного приложения FatSecret мы также установили, что среднее количество потребляемых калорий в сутки для 10 случайно отобранных мужчин составляет 2320,2 (SD = 388), а для 10 случайно отобранных женщин – 1992,5 (SD = 323). Опираясь на данную информацию, мы можем более точно рассчитать величину эффекта. Для этого

необходимо нажать на Determine/Определить рядом с Effect size d. Учитывая равное число участников исследования в двух группах, вводим необходимые параметры в открывшееся диалоговое окно: Mean / Среднее значение и стандартное отклонение / SD для group 1 / Группы 1 (мужчины) и group 2 / Группы 2 (женщины) (рис. 5). Нажимаем Calculate/Рассчитать и получаем величину эффекта Effect size d равную 0,9188. При нажатии Calculate and transfer to main window / Рассчитать и перенести в главное окно величина эффекта рассчитывается и автоматически переносится в графу Effect size d основного диалогового окна программы.

После нажатия Calculate/Рассчитать в правом нижнем углу основного диалогового окна

Input Parameters		Output Parameters	
Determine =>	Effect size d: 0.5	Noncentrality parameter δ	?
	α err prob: 0.05	Critical t	?
	Power (1- β err prob): 0.80	Df	?
	Allocation ratio N2/N1: 1	Sample size group 1	?
		Sample size group 2	?
		Total sample size	?
		Actual power	?

Рис. 3. Ввод необходимых параметров (двусторонний тест, величина эффекта, α -ошибка, мощность, соотношение участников в двух группах) для расчета объема выборки

Fig. 3. Input parameters (two-tailed test, effect size, alpha-error, power, allocation ratio) for sample size calculation

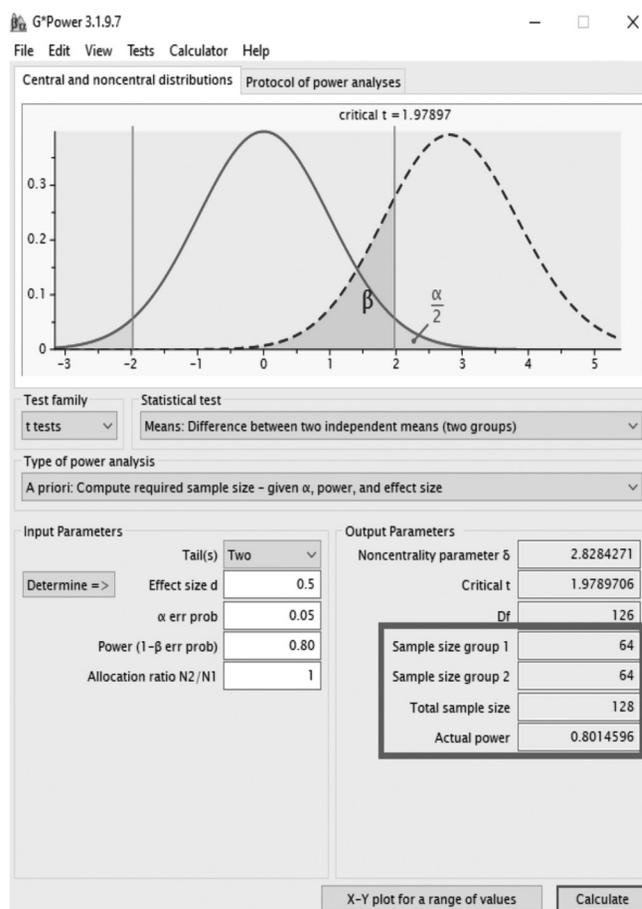


Рис. 4. Результат расчета необходимого объема выборки для статистического теста при сравнении двух независимых средних значений (объяснение в тексте)

Fig. 4. G*Power output for sample size calculation for comparing two independent means (see text for details)

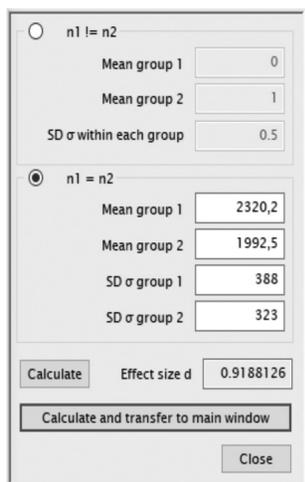


Рис. 5. Диалоговое окно для расчета Effect size d / Величины эффекта

Fig. 5. G*Power dialog box for effect size calculation

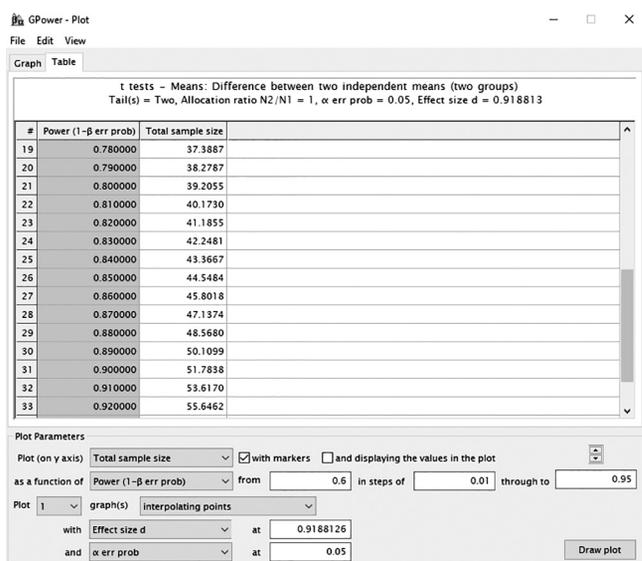


Рис. 7. Зависимость размера выборки Total sample size от мощности Power (1-β err prob), представленная в виде числовых значений

Fig. 7. Numeric association between total sample size and statistical power

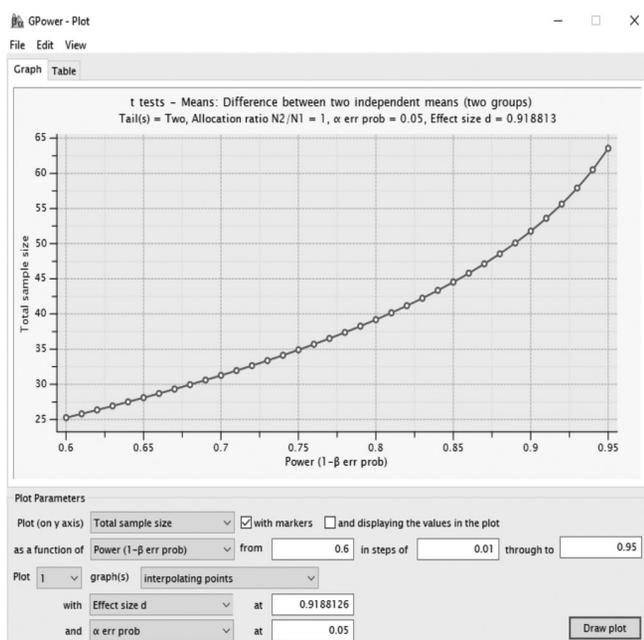


Рис. 6. Зависимость размера выборки Total sample size от мощности Power (1-β err prob) представленная в виде графика

Fig. 6. Graphical presentation of the association between total sample size and statistical power

программы вновь получаем результат для большей величины эффекта при неизменных остальных параметрах. Исходя из заданных нами параметров, для изучения суточного количества потребляемых калорий лицами разного пола нам необходимо включить в исследование минимум 40 человек, 20 мужчин и 20 женщин. Можно заметить, что для определе-

ния большей величины эффекта (при прочих равных параметрах) потребуется меньшее количество участников исследования.

При нажатии на X-Y plot for a range of values/ График X-Y для диапазона значений программа позволяет построить графики для одного из следующих параметров: α-ошибка, величина эффекта, мощность и объем выборки, в зависимости от диапазона значений остальных параметров. Выбрав Total sample size / Общий объем выборки в графе Plot (on y axis) как функции от Power (1-β err prob)/ мощность в диапазоне значений от 0,6 до 0,95 с шагом 0,01, мы видим, как меняется мощность исследования в зависимости от увеличения объема выборки с учетом заданных параметров величины эффекта (Effect size d) и α-ошибки (α err prob) (рис. 6). При выборе раздела Table/Таблица данные, представленные графически, трансформируются в таблицу, также представляющую зависимость мощности от размера выборки. Например, включение в исследование 52 человек позволит увеличить мощность исследования до 90 % (рис. 7).

Также возможно построение графика зависимости размера выборки Total sample size от мощности Power (1-β err prob) с учетом величины эффекта Effect size d, что может быть удобно в том случае, когда величина эффекта определяется априори. График еще раз наглядно

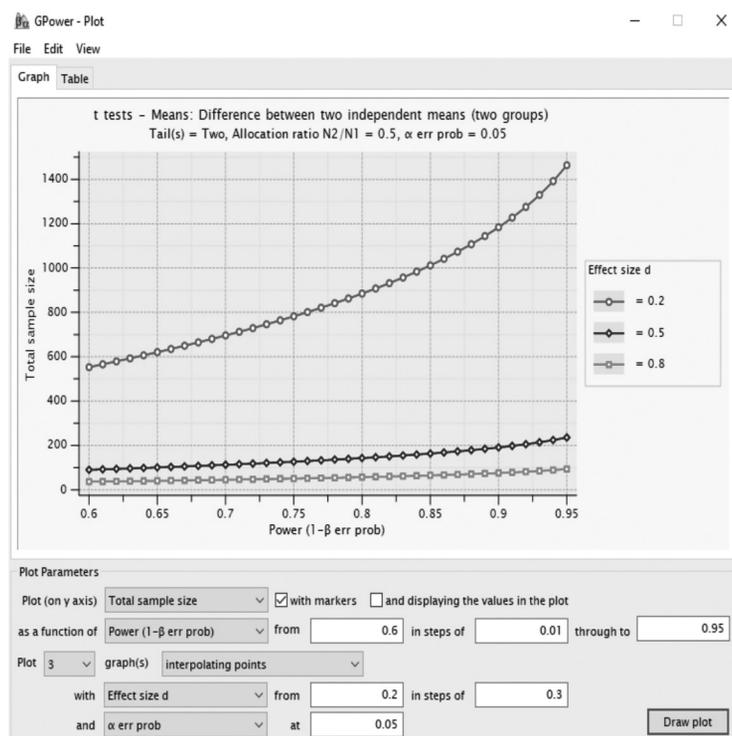


Рис. 8. Зависимость размера выборки Total sample size от мощности Power ($1-\beta$ err prob) с учетом величины эффекта Effect size d

Fig. 8. Visualization of the associations between total sample size and statistical power by effect size

демонстрирует, что для выявления небольшой величины эффекта требуется значительное увеличение объема выборки при равенстве значений прочих параметров (рис. 8).

Расчет объема выборки при использовании однофакторного дисперсионного анализа

Далее разберем на конкретном примере расчет необходимого объема выборки для сравнения количественных показателей, подчиняющихся закону нормального распределения, при наличии трех и более групп. Для данного анализа в наибольшей степени подходит одномерный (однофакторный) дисперсионный анализ для независимых групп (One-way ANalysis Of VAriance, ANOVA) [4].

В данном случае пример формулировки H_0 и H_1 может быть записан следующим образом:

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$, популяционные средние значения в k изучаемых группах не различаются ($k \geq 3$);

$H_1: \mu_1 \neq \mu_2 \neq \dots = \mu_k$, популяционные средние значения не равны.

Тестирование H_0 проводится путем сравнения дисперсии изучаемого признака между группами и внутри групп с использованием критерия Фишера (F-тест). При равенстве межгрупповой и внутригрупповой дисперсий принимается H_0 об отсутствии межгрупповых

различий между средними. В том случае, если межгрупповая дисперсия больше внутригрупповой дисперсии, H_0 отклоняется и принимается H_1 , свидетельствующая о том, что средние значения между группами различаются. Поэтому для расчета объема выборки в графе Test family / Семейство тестов из выпадающего списка выбираем категорию F-тест. Далее в графе Statistical test / Статистический тест из выпадающего списка нужно выбрать ANOVA: Fixed effects, omnibus, one-way. Как и ранее, в графе Type of power analysis/тип анализа мощности выбираем A priori: Compute required sample size – given α , power, and effect size / Априори: вычислить необходимый размер выборки — учитывая величину α -ошибки, мощность и величину эффекта. Задаем параметры α -ошибки (α err prob) равной 0,05 и мощности Power ($1-\beta$ err prob) равной 0,80, как это описывалось ранее. Указываем количество сравниваемых групп в соответствующей графе Number of groups / Количество групп. Значение величины эффекта Effect size d можно указать, если оно заранее известно, или рассчитать, нажав на Determine/Определить. Рассмотрим второй вариант на конкретном примере. Предположим, нас интересует, на сколько килограммов в среднем снижается масса тела людей, соблюдающих разные виды

диеты в течение 6 месяцев. Пилотное исследование по изучению эффективности 4 видов диет (по 10 участников в каждой группе) показало, что при интервальном голодании участники похудели в среднем на 4,36 кг (SD = 1,93), при соблюдении кетодиеты на 2,82 кг (SD = 1,57), при сиртфуд-диете на 5,94 (SD = 2,14), при палеодиете – на 6,22 (SD = 2,44).

Для расчета величины эффекта нам требуется объединенное стандартное отклонение (SD σ within each group), которое можно рассчитать по формуле [5]:

$$SD = \sqrt{\frac{(n_1-1) \cdot SD_1^2 + (n_2-1) \cdot SD_2^2 + (n_3-1) \cdot SD_3^2 + (n_4-1) \cdot SD_4^2}{(n_1-1) + (n_2-1) + (n_3-1) + (n_4-1) - 4}}$$

где n – количество участников в каждой группе, SD – стандартное отклонение для каждой группы.

Подставив в формулу соответствующие значения получаем объединенное стандартное отклонение равное 2,23.

Для расчета величины эффекта в диалоговом окне, открывшемся при нажатии на Determine/Определить, в графе Select procedure/Выбор процедуры оставляем Effect size from means/ Величина эффекта, исходя из средних значений, далее указываем количество групп в графе Number of groups / Количество групп и объединенное стандартное отклонение в графе SD σ within each group (рис. 9).

После нажатия кнопки Calculate/Рассчитать получаем размер величины эффекта, равный

0,61, и копируем его в основное диалоговое окно или нажимаем Calculate and transfer to main window для автоматического заполнения соответствующего поля.

После нажатия на Calculate/Рассчитать в разделе Output parameters / Выходные параметры основного диалогового окна для данного статистического теста мы видим, что в данное исследование необходимо включить минимум 36 человек (Total sample size) (рис. 10).

При нажатии на X-Y plot for a range of values / График X-Y для диапазона значений также можно построить графики и таблицы зависимости размера выборки от мощности с учетом величины эффекта, как это рассматривалось выше.

Расчет объема выборки для применения критерия Вилкоксона для одной выборки

Когда необходимо рассчитать объем выборки для сравнения значений в исследуемой группе с каким-то одним заданным значением, и рас-

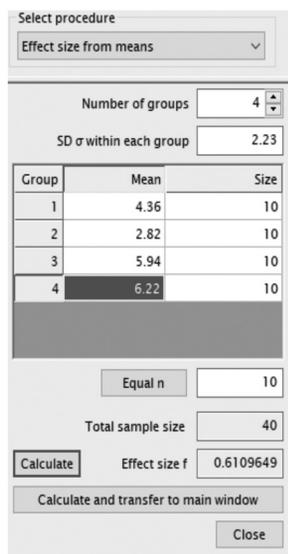


Рис. 9. Диалоговое окно для расчета Effect size d / Величины эффекта
 Fig. 9. Dialog box for effect size calculation

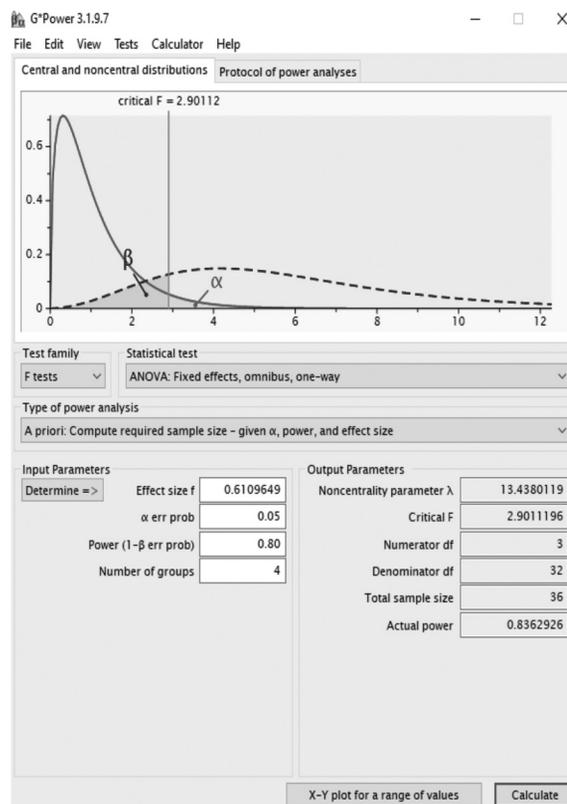


Рис. 10. Результат расчета необходимого объема выборки для статистического теста ANOVA: Fixed effects, omnibus, one-way (объяснение в тексте)
 Fig. 10. Results of sample size calculation for fixed-effects one-way ANOVA (see text for details)

пределение признака отличается от нормального, используется ранговый критерий Вилкоксона для одной выборки. Для этого в графе Test family / Семейство тестов в выпадающем окне необходимо выбрать *t* tests / *t*-критерии, а затем в графе Statistical test / Статистический тест выбрать Wilcoxon signed-rank test (one sample case) / Ранговый тест Вилкоксона (одна выборка) (рис. 11). Для расчета объема выборки, который проводится перед началом исследования, открываем выпадающий список в графе Type of power analysis / Тип анализа мощности и выбираем A priori: Compute required sample size – given α , power, and effect size / Априори: вычислить необходимый размер выборки — учитывая величину α -ошибки, мощность и величину эффекта. Далее в разделе Input parameters / Входные параметры необходимо задать 5 параметров: Tail(s), Parent distribution, Effect size *d*, α err prob, Power ($1-\beta$ err prob).

Ранее мы упомянули, что G*Power использует принцип тестирования нулевой гипотезы Неймана–Пирсона [1], формулировка которой определяет выбор одностороннего (one-tailed) или двухстороннего критерия (two-tailed) в графе Tail(s). Аналогично параметрическим критериям, при непараметрических критериях мы будем использовать односторонние критерии только в том случае, если точно знаем направленность изучаемой взаимосвязи. Затем нам необходимо выбрать пункт в графе Parent distribution / Теоретическое распределение. Выпадающий список предлагает варианты Laplace, Logistic, или min ARE (minimal asymptotic relative efficiency), выбрать которые необходимо в зависимости от предполагаемого распределения переменной. Опция min ARE подходит по умолчанию, если существуют сомнения о типе распределения, или оно вообще неизвестно. Подробное обоснование идеи и принципов расчета при выборе разных видов родительского распределения представлены в руководстве по использованию G*Power [6]. Параметр Effect size *d* / Величина эффекта задается в зависимости от величины различий в соответствии с целью исследования и предположения о клинической значимости эффекта. Значение величины эффекта определяется аналогичным способом, который был рассмотрен выше при описании параметрических тестов. В двух последних графах задается

значение α -ошибки (α err prob) и необходимая мощность исследования Power ($1-\beta$ err prob).

Рассмотрим расчет объема выборки, необходимой при сравнении центральных тенденций для признаков, распределение которой не подчиняется закону нормального распределения, с неким заданным конкретным значением. Например, данный подход будет применим в ситуации, когда мы хотим проверить, является ли среднее количество детей в семьях Архангельска больше 1? В нашем примере нулевая и альтернативная гипотезы будут сформулированы следующим образом:

$H_0: \mu = 1$, то есть среднее число детей в семьях Архангельска равно 1; $H_1: \mu > 1$, то есть среднее число детей в семьях Архангельска больше 1.

После этого следует заполнить раздел Input parameters / Входные параметры. В нашем примере мы планируем рассчитать объем выборки для одностороннего теста, потому что нас интересуют только значения большие, чем это определено нулевой гипотезой (1 ребенок). Поскольку у нас нет справочной информации про распределение, в родительском распреде-

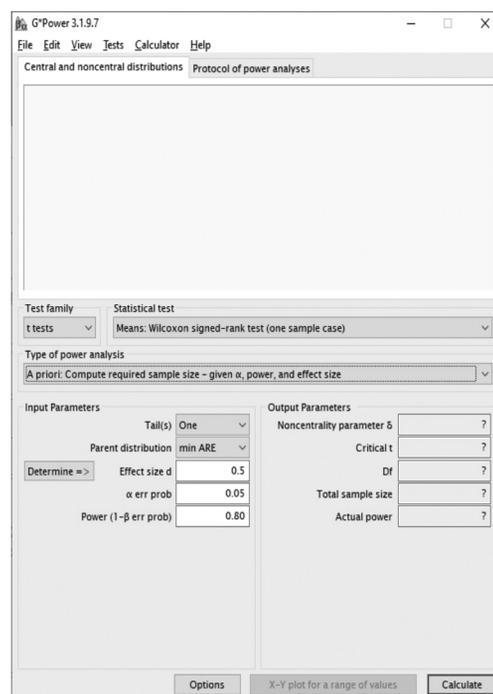


Рис. 11. Интерфейс G*Power при выборе одновыборочного критерия Вилкоксона
Fig. 11. G*Power dialog box for one-sample Wilcoxon signed-rank test

лении мы выберем min ARE. Уровень α -ошибки установим 0,05 и мощность 0,80. Допустим, мы планируем определить большую величину эффекта, поэтому в графу Effect size d вводим величину 0,8 (рис. 12).

После нажатия на Calculate/Рассчитать мы получим результат (рис. 13), который появится

в разделе Output parameters / Выходные параметры. Исходя из расчетов, можно сделать вывод о том, что в данное исследование необходимо включить минимум 13 семей Архангельска (Total sample size).

Допустим, что результаты ранее проведенных исследований показали, что количество

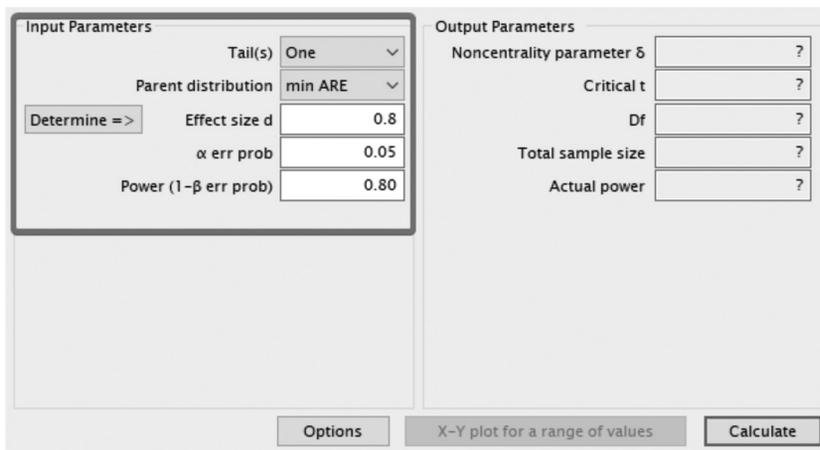


Рис. 12. Ввод необходимых параметров для расчета объема выборки для одновыборочного критерия Вилкоксона
Fig. 12. Input parameters for one-sample Wilcoxon signed-rank test sample size calculation

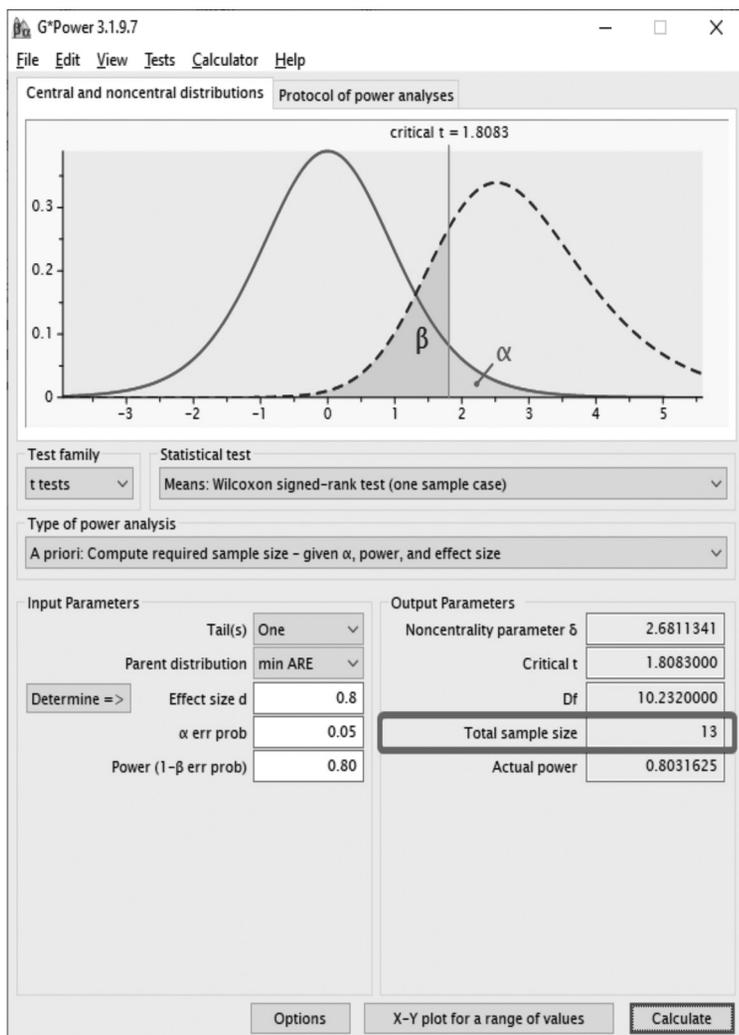


Рис. 13. Результат расчета необходимого объема выборки для одновыборочного критерия Вилкоксона
Fig. 13. Results of sample size calculation for one-sample Wilcoxon signed-rank test

детей на семью в Архангельске в среднем составляет 1,24 (стандартное отклонение 0,49). Данная информация позволит нам более точно рассчитать величину эффекта. Для этого выберем опцию Determine/Определить напротив

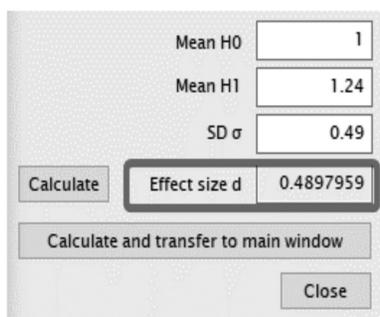


Рис. 14. Диалоговое окно для расчета Effect size d / Величины эффекта

Fig. 14. Dialog box for effect size calculation

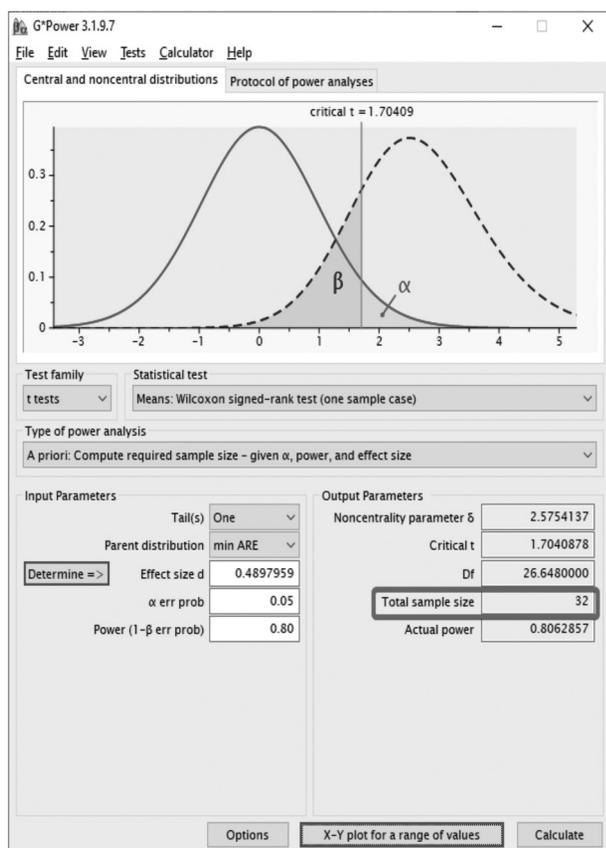


Рис. 15. Результат перерасчета размера выборки при самостоятельном определении размера эффекта (Effect size d) для одновыборочного критерия Вилкоксона

Fig. 15. Results of sample size calculation for one-sample Wilcoxon signed-rank test with defined effect size

параметра Effect size d. В появившемся окне нам необходимо ввести ранее найденные Mean / Среднее значение в графу Mean H1 и SD / Стандартное отклонение, а в графе Mean H0 указать значение 1 согласно сформулированной гипотезе (рис. 14). Нажав Calculate/Рассчитать, мы получим величину эффекта Effect size d равную 0,4897959. При помощи Calculate and transfer to main window / Рассчитать и перенести в главное окно величина эффекта автоматически переносится в графу Effect size d в главном окне программы.

Поскольку мы рассчитали новую величину эффекта, необходимо произвести перерасчет объема выборки. Для этого снова нажмем Calculate/Рассчитать. Теперь с учетом меньшей величины эффекта и сохранения остальных параметров нам потребуется включить в исследование 32 семьи Архангельска для ответа на поставленный исследовательский вопрос (рис. 15).

Кроме того, мы можем нажать X-Y plot for a range of values/ График X-Y для диапазона значений и в появившемся окне выбрать Draw plot / Построить график, чтобы построить график зависимости интересующих нас параметров: α-ошибки, величины эффекта, мощности или объема выборки. Так, например, можно наглядно посмотреть, как меняется мощность исследования при изменении объема выборки, а во вкладке Table/Таблица можно узнать, что для достижения 90 % мощности исследования необходимо включить в исследование 43 семьи (рис. 16, 17).

Расчет объема выборки для U-теста Манна-Уитни для двух независимых выборок

Далее рассмотрим ситуацию, при которой необходимо рассчитать объем выборки, когда мы хотим проверить, отличаются ли центральные тенденции между двумя независимыми группами, если распределение признака в этих группах отличается от нормального. В этом случае методом выбора для расчета будет U-тест Манна-Уитни для двух независимых выборок. Для этого в графе Test family / Семейство тестов выбираем t tests / t-тесты, затем Wilcoxon-Mann-Whitney test (two groups) / Тест Вилкоксона-Манна-Уитни (две группы) в графе Statistical test / Статистический тест.

Предположим, что мы разработали новую методику послеоперационного ухода. Нам необходимо проверить, различаются ли средние оценки

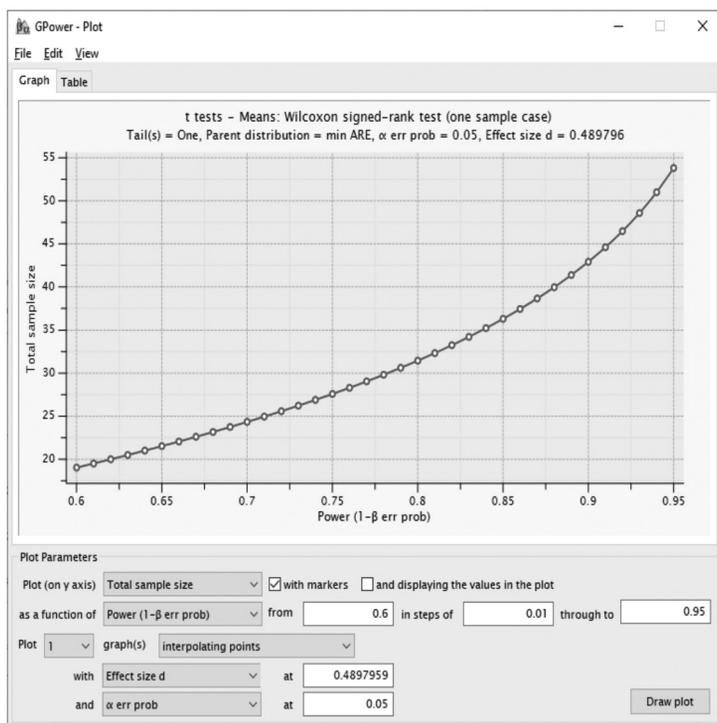


Рис. 16. Зависимость размера выборки Total sample size от мощности Power (1-β err prob), представленная в виде графика
Fig. 16. Graphical presentation of the association between total sample size and statistical power

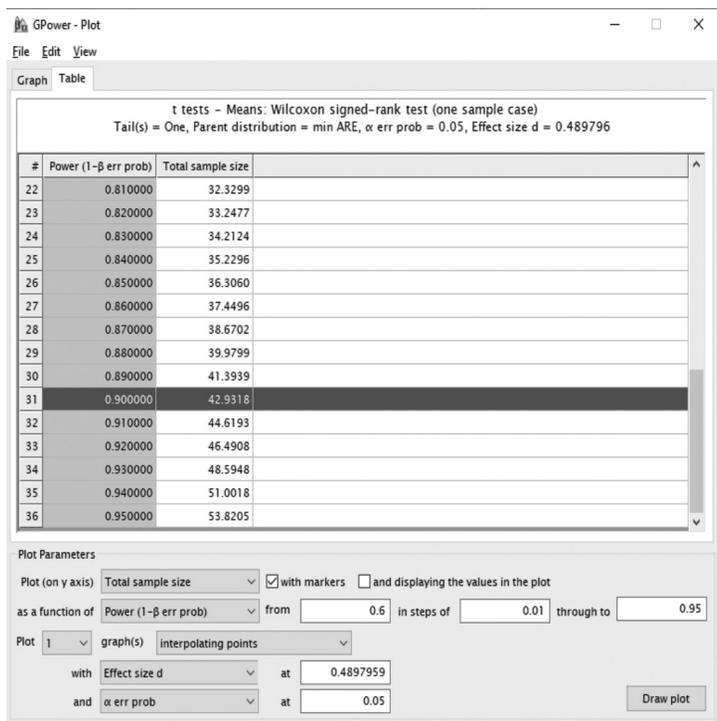


Рис. 17. Зависимость размера выборки Total sample size от мощности Power (1-β err prob) для одновыборочного критерия Вилкоксона, представленная в виде числовых значений
Fig. 17. Numeric association between total sample size and statistical power for one-sample Wilcoxon test

по шкале боли в двух группах лечения, использующих традиционную и новую методики послеоперационного ухода. В этом случае сформулированные гипотезы могут выглядеть так:

H0: $\mu_1 = \mu_2$, средняя оценка по шкале боли в изучаемых группах не различается;

H1: $\mu_1 \neq \mu_2$, средняя оценка по шкале боли в изучаемых группах различается.

В графе Type of power analysis / Тип анализа мощности выбираем A priori: Compute required sample size – given α, power, and effect size / Априори: вычислить необходимый размер вы-

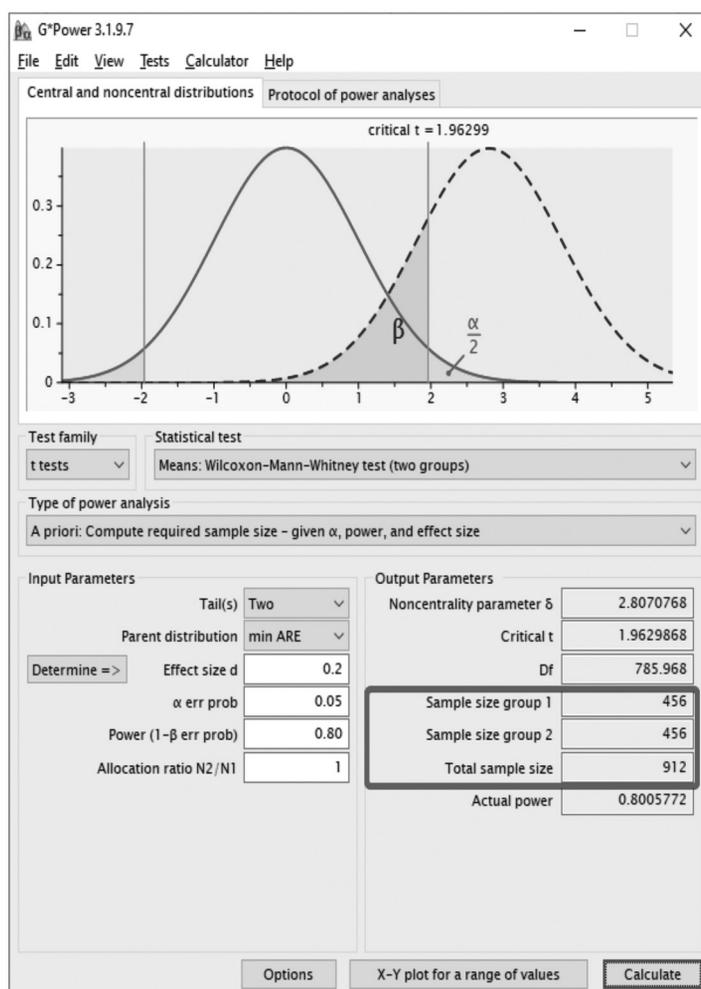


Рис. 18. Результаты расчета объема выборки для критерия Манна–Уитни для двух независимых групп

Fig. 18. Results of sample size calculation for Mann-Whitney test for two independent samples

борки, учитывая величину α -ошибки, мощность и величину эффекта. На основании описанного ранее, задаем следующие параметры: двусторонний тест (two-tailed test), теоретическое распределение (Parent distribution) как min ARE, α -ошибку (α err prob) указываем равной 0,05 и мощность Power (1- β err prob) равной 0,80. Допустим, что мы не имеем информации о размере эффекта, и предположим, что его величина мала (Effect size d = 0,2). Еще одним параметром, который необходимо указать, станет Allocation ratio N2/N1 / Коэффициент распределения N2/N1, отражающий соотношение участников исследования в сравниваемых группах. Оставим его значение как 1 (группы равны). После ввода параметров нажимаем Calculate/Рассчитать и получаем результаты. Объем выборки составит 912 человек, из которых 456 войдут в группу лечения по традиционной методике и 456 – по новой методике (рис. 18).

Допустим, что мы провели пилотное исследование и обнаружили, что при новой методике лечения средняя оценка по шкале боли составила 4,45 (Mean group 1), а стандартное отклонение 2,49 (SD group 1). В группе лечения по традиционной методике соответствующие значения составили 6,08 (Mean group 2) и 2,56 (SD Group 2). Исходя из собранных данных, мы можем самостоятельно рассчитать более точную величину эффекта, подставив эти значения во всплывающем окне после нажатия Determine/ Определить напротив параметра Effect size d, а затем произвести перерасчет объема выборки. При таких условиях величина эффекта оказывается ниже, что десятикратно снижает общий необходимый объем выборки (90 человека) при условии сохранения других параметров (рис. 19).

Когда нам необходимо рассчитать объем выборки, чтобы проверить отличаются ли средние значения в трех и более независимых группах,

Рис. 19. Результат перерасчета размера выборки при самостоятельном определении размера эффекта (Effect size d) для критерия Манна-Уитни

Fig. 19. Results of sample size calculation for Mann-Whitney test with defined effect size

а распределение признака в этих группах не подчиняется закону нормального распределения, применяют тест Краскела–Уоллиса, который является непараметрическим аналогом одномерного (однофакторного) дисперсионного анализа (One-way ANalysis Of VAriance, ANOVA) [4]. G*Power не дает нам стандартного способа вычисления объема выборки для этого непараметрического теста. В таком случае можно воспользоваться эмпирическим правилом: если мы планируем использовать непараметрический тест, следует рассчитать объем выборки, необходимой для соответствующего параметрического теста, и добавить к ней 15 % [8].

Заключение. Расчет объема выборки в зависимости от цели и задач исследования на этапе планирования позволяет, с одной стороны, избежать излишних затрат ресурсов, с другой – добиться необходимой мощности ис-

следования. Для грамотного расчета объема выборки необходимо не просто владеть методиками использования программного обеспечения, но и хорошо представлять изучаемое явление, понимать, какие различия между сравниваемыми группами можно считать клинически значимыми. Стоит также заметить, что рассчитанный с помощью программного обеспечения показатель — это не руководство к действию, а лишь то минимальное число участников исследования, которых мы должны включить в статистический анализ для достижения необходимых параметров (мощность исследования, величина эффекта и пр.). Кроме этого, необходимо обязательно учитывать дизайн исследования, уровень отклика потенциальных участников, возможность их выбывания из продольных исследований, то есть рассчитанный размер выборки целесообразно увеличить минимум на 15 %.

Сведения об авторах:

Кригер Екатерина Анатольевна — кандидат медицинских наук, доцент, научный сотрудник международного центра научных компетенций центральной научно-исследовательской лаборатории федерального государственного бюджетного образовательного учреждения высшего образования «Северный государственный медицинский университет» (Архангельск) Минздрава России; 163069, г. Архангельск, Троицкий проспект, д. 51; e-mail: kate-krieger@mail.ru; ORCID 0000-0001-5179-5737; SPIN: 2686-7226

Драчев Сергей Николаевич — доктор философии в области медицинских наук, кандидат медицинских наук, магистр общественного здоровья, доцент кафедры ортопедической стоматологии федерального государственного бюджетного образовательного учреждения высшего образования «Северный государственный медицинский университет» (Архангельск) Минздрава России; 163069, г. Архангельск, Троицкий проспект, д. 51; e-mail: drachevsn@mail.ru; ORCID 0000-0002-1548-690X; SPIN: 3879-8612.

Митькин Никита Андреевич — лаборант-исследователь международного центра научных компетенций Центральной научно-исследовательской лаборатории, ассистент кафедры методологии научных исследований федерального государственного бюджетного образовательного учреждения высшего образования «Северный государственный медицинский универ-

ситет» (Архангельск) Минздрава России; 163069, г. Архангельск, Троицкий проспект, д. 51; e-mail: n.a.mitkin@gmail.com; ORCID 0000-0002-0027-8155

Постоев Виталий Александрович — кандидат медицинских наук, и. о. директора Института развития научных и проектных компетенций. Заведующий Архангельской международной школой общественного здоровья, федерального государственного бюджетного образовательного учреждения высшего образования «Северный государственный медицинский университет» (Архангельск) Минздрава России; 163069, г. Архангельск, Троицкий проспект, д. 51; e-mail: ispha@nsmu.ru; ORCID: 0000-0003-4982-4169

Гржибовский Андрей Мечиславович — доктор медицинских наук, начальник управления научно-инновационной работы, заведующий центральной научно-исследовательской лабораторией федерального государственного бюджетного образовательного учреждения высшего образования Северный государственный медицинский университет (Архангельск) Минздрава России; 163069, г. Архангельск, Троицкий проспект, д. 51; e-mail: A.Grjibovski@yandex.ru; ORCID 0000-0002-5464-0498; SPIN: 5118-0081.

Information about the authors:

Ekaterina A. Krieger — Cand. of Sci. (Med.), Associate Professor, Research fellow of the International Research Competence Centre, Central Scientific Research Laboratory, Northern State Medical University; 51 Troitskiy Av., Arkhangelsk, 163069, Russia; e-mail: kate-krieger@mail.ru; ORCID 0000-0001-5179-5737; SPIN: 2686-7226.

Sergei N. Drachev — Cand. of Sci. (Med), MPH, Associate Professor at the Department of Prosthodontics, Northern State Medical University; 51 Troitskiy Av., Arkhangelsk, 163069, Russia; e-mail: drachevsn@mail.ru; ORCID 0000-0002-1548-690X; SPIN: 3879-8612.

Nikita A. Mitkin — Junior researcher, International Research Competence, Central Scientific Research Laboratory, Assistant of the Department of Scientific Research Methodology, Northern State Medical University; 51 Troitskiy Av., Arkhangelsk, 163069, Russia; e-mail: n.a.mitkin@gmail.com; ORCID 0000-0002-0027-8155

Vitaly A. Postoev — Cand. of Sci (Med.), Acting Head of the Department of Research Methodology, Head of Arkhangelsk School of Public Health, Northern State Medical University, Troitskiy Ave., 51, Arkhangelsk, 163069, Russia; e-mail: ispha@nsmu.ru; ORCID: 0000-0003-4982-4169

Andrej M. Grjibovski — Doctor of Medicine, Head of the Directorate for Research and Innovations, Director of the Central Scientific Research Laboratory, Northern State Medical University; 51 Troitskiy Av., Arkhangelsk, 163069, Russia; e-mail: A.Grjibovski@yandex.ru; ORCID 0000-0002-5464-0498; SPIN: 5118-0081.

Вклад авторов. Все авторы подтверждают соответствие своего авторства, согласно международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией).

Author contribution. All authors according to the ICMJE criteria participated in the development of the concept of the article, obtaining and analyzing factual data, writing and editing the text of the article, checking and approving the text of the article.

Потенциальный конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Disclosure. The authors declare that they have no competing interests.

Поступила/Received: 04.06.2023

Принята к печати/Accepted: 10.06.2023

Опубликована/Published: 30.06.2023

ЛИТЕРАТУРА / REFERENCES

1. Kadam P., Bhalerao S. Sample size calculation. *Int J Ayurveda Res.* 2010, Vol. 1, № 1, pp. 55–57. doi: 10.4103/0974-7788.59946
2. Chan Y.H. Biostatistics 102: Quantitative Data – Parametric & Non-parametric Tests. *Singapore Med J*, 2003, Vol. 44, № 8, 391–396.
3. Kim H.Y. Analysis of variance (ANOVA) comparing means of more than two groups. *Restor Dent Endod.*, 2014, Vol. 39, № 1, pp. 74–77. doi: 10.5395/rde.2014.39.1.74.
4. Sainani KL. Dealing with non-normal data. *PM R*, 2012, Vol. 12, № 4, pp. 1001–1005. doi: 10.1016/j.pmrj.2012.10.013
5. Kang H. Sample size determination and power analysis using the G*Power software. *J Educ Eval Health Prof.*, 2021, Vol. 18, P. 17. doi: 10.3352/jeehp.2021.18.17
6. Faul F., Erdfelder E., Lang A.G., Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.*, 2007, Vol. 39, № 2, pp.175–191. doi: 10.3758/bf03193146
7. Гржибовский А.М., Гвоздецкий А.Н. Интерпретация величины р и альтернативы ее использованию в биомедицинских исследованиях // *Экология человека*. 2022. Т. 29, № 3. С. 67–76 [Grjibovski AM, Gvozdeckii AN. Interpretation of and alternatives to p-values in biomedical sciences. *Ekologiya cheloveka (Human Ecology)*, 2022, Vol. 29, № 3, 67–76 (In Russ.)]. doi: <https://doi.org/10.17816/humeco97249>.
8. Lehmann E.L., D'Abbrera H.J.M. *Nonparametrics: Statistical Methods Based on Ranks*: Holden-Day; 1975, p. 457.