Статистическая мощность эксперимента в Excel

Недавно прочитал книгу Майкла Херцога с соавторами [Статистика и планирование эксперимента для непосвященных](). В ней я в очередной раз встретился со *статистической мощностью*. До сих пор я относился к этой статистике несколько академически. Что называется, не чувствовал её на кончиках пальцев. Некоторые разделы книги Херцога меня заинтересовали, я построил пару моделей в Excel, и до меня дошла суть статистической мощности. Я понял, что статистическая мощность – вероятность получить в эксперименте статистически значимые результаты. Кроме того, я нашел формулу расчета статистической мощности в Excel.

=НОРМ.ОБР(СЛМАССИВ(20;;0;1;ЛОЖЬ);0;1)

	A	B	C	D
1	Выборка 1	Выборка 2		
2	-0,9041	1,7808		
3	-0,4602	0,1226		
4	0,0442	0,4006		
5	-0,2473	0,2032		
6	-1,4949	0,8501		
7	0,0448	-0,4281		
8	0,1958	0,6884		
9	1,6127	0,2816		
10	-0,3526	3,0480		
11	0,9070	1,8653		
12	1,8746	1,7059		
13	1,0947	2,5406		
14	-0,7834	2,0201		
15	0,6775	1,0512		
16	1,4339	0,8067		
17	1,9777	2,2991		
18	-1,2532	0,1260		
19	0,2287	0,4691		
20	2,6330	1,4602		
21	-0,0666	0,6366		
22				

Рис. 1. Нормально распределенные случайные величины. Выборка 1 – N(0;1), Выборка 2 – N(1;1)

Если вы не знакомы со статистическим выводом, рекомендую начать с [t-статистики Стьюдента]().

Суть процедуры статистического вывода

Допустим, мы извлекаем две выборки по 20 элементов в каждой из двух генеральных совокупностей. Такая схема эксперимента называется двухвыборочной. Наша нулевая гипотеза H_0 состоит в том, что математические ожидания генеральных совокупностей равны $\mu_1 = \mu_2$. В качестве альтернативной выберем самую общую гипотезу $\mu_1 \neq \mu_2$.

Но мы не знаем истинные значения μ_1 и μ_2. В нашем распоряжении только выборочные средние значения $\bar{x}_1$ и $\bar{x}_2$. Причем они не равны. Задача процедуры статистического вывода – ответить на вопрос: есть ли основания при заданном уровне значимости отвергнуть нулевую гипотезу? Т.е., можно ли объяснить различия между средними значениями $\bar{x}_1$ и $\bar{x}_2$ естественной вариабельностью при отборе значений из одной генеральной совокупности ($\mu_1 = \mu_2$). Или это настолько маловероятно, что различия между $\bar{x}_1$ и $\bar{x}_2$ скорее связаны с тем, что они извлечены из разных генеральных совокупностей ($\mu_1 \neq \mu_2$).

Модель в Excel

Рассмотрим модель в Excel (см. рис. 1), в которой мы генерим нормально распределенные случайные величины из двух генеральных совокупностей: N(0;1) и N(1;1). Здесь N(0;1) – нормальное распределение с математическим ожиданием $\mu_1 = 0$ и $\sigma_1 = 1$. N(1;1) – нормальное распределение с математическим ожиданием $\mu_2 = 1$ и $\sigma_2 = 1$.

Для дальнейшей иллюстрации примера я сохранил сгенерированные формулами случайные числа, как значения. В приложенном Excel-файле на листе «Рис. 1» можно найти формулы, дающие случайные числа. Эти значения волатильны. Они изменяются после любого пересчета в Excel.

Двухвыборочный t-тест в Пакете анализа

В Excel есть средство, которое позволяет выполнить двухвыборочный t-тест. Пройдите *Данные –> Анализ данных* (кнопка расположена на правом краю ленты). Выберите *Двухвыборочный t-тест с одинаковыми дисперсиями* или *Двухвыборочный t-тест с различными дисперсиями*. Первый вид теста выполняется, если у нас есть основания предполагать, что неизвестные дисперсии двух выборок одинаковы.

	A	B	C	D	E	F
1	Выборка 1	Выборка 2		Двухвыборочный t-тест с одинаковыми дисперсиями		
2	-0,9041	1,7808				
3	-0,4602	0,1226			*Переменная 1*	*Переменная 2*
4	0,0442	0,4006		Среднее	0,3581	1,0964
5	-0,2473	0,2032		Дисперсия	1,2980	0,8861
6	-1,4949	0,8501		Наблюдения	20	20
7	0,0448	-0,4281		Объединенная дисперсия	1,0921	
8	0,1958	0,6884		Гипотетическая разность средних	0	
9	1,6127	0,2816		df	38	
10	-0,3526	3,0480		t-статистика	-2,2341	
11	0,9070	1,8653		P(T<=t) одностороннее	0,0157	
12	1,8746	1,7059		t критическое одностороннее	1,6860	
13	1,0947	2,5406		P(T<=t) двухстороннее	0,0314	
14	-0,7834	2,0201		t критическое двухстороннее	2,0244	
15	0,6775	1,0512				
16	1,4339	0,8067				
17	1,9777	2,2991				
18	-1,2532	0,1260				
19	0,2287	0,4691				
20	2,6330	1,4602				
21	-0,0666	0,6366				

Рис. 2. Двухвыборочный t-тест с одинаковыми дисперсиями (надстройка *Пакет анализа*)

Если *Анализ данных* у вас на ленте не отображается, пройдите *Файл –> Параметры*. В окне *Параметры Excel* перейдите на вкладку *Надстройки*. В нижней части окна в поле *Управление* выберите *Надстройки Excel*. Кликните на кнопке *Перейти*. Поставьте галочку напротив *Пакет анализа*.

Двухвыборочный t-тест с помощью формул в Excel

В учебных целях такой подход я не могу рекомендовать, поскольку в нём все вычисления «находятся под капотом». Выполним шаги последовательно:

	A	B	C	D	E	F	G
1	Выборка 1	Выборка 2					
2	-0,9041	1,7808		Выборочное среднее 1	$\bar{x}_1$	0,3581	=СРЗНАЧ(A2:A21)
3	-0,4602	0,1226		Выборочное среднее 2	$\bar{x}_2$	1,0964	=СРЗНАЧ(B2:B21)
4	0,0442	0,4006		Число элементов 1	n_1	20	=СЧЁТЗ(A2:A21)
5	-0,2473	0,2032		Число элементов 2	n_2	20	=СЧЁТЗ(B2:B21)
6	-1,4949	0,8501		Дисперсия 1	σ_1^2	1,2980	=ДИСП.В(A2:A21)
7	0,0448	-0,4281		Дисперсия 2	σ_2^2	0,8861	=ДИСП.В(B2:B21)
8	0,1958	0,6884		Объединенная дисперсия	σ^2	1,0921	=СРЗНАЧ(F6:F7)
9	1,6127	0,2816		Выборочное стандартное отклонение	s	1,0450	=КОРЕНЬ(F8)
10	-0,3526	3,0480		Число степеней свободы	df	38	=F4+F5-2
11	0,9070	1,8653		Размер эффекта	d	0,7065	=ABS(F2-F3)/F9
12	1,8746	1,7059		t-статистика	t	2,2341	=F11*КОРЕНЬ(F4/2)
13	1,0947	2,5406		Вероятность ошибки I-го рода	α	0,05	
14	-0,7834	2,0201		Критическое односторонне t	t_α	1,6860	=СТЬЮДЕНТ.ОБР(1-F13;F10)
15	0,6775	1,0512		Критическое двусторонне t	t_α	2,0244	=СТЬЮДЕНТ.ОБР.2Х(F13;F10)
16	1,4339	0,8067		Вероятность p ($t \le t_\alpha$) одностороннее	p	0,0157	=1-СТЬЮДЕНТ.РАСП(F12;F10;ИСТИНА)
17	1,9777	2,2991		Вероятность p ($t \le t_\alpha$) двустороннее	p	0,0314	=СТЬЮДЕНТ.РАСП.2Х(F12;F10)
18	-1,2532	0,1260					
19	0,2287	0,4691					
20	2,6330	1,4602					
21	-0,0666	0,6366					

Рис. 3. Двухвыборочный t-тест с одинаковыми дисперсиями (формулы)

Ошибка I-го рода

Выше мы сказали, что в качестве альтернативной рассматриваем самую общую гипотезу $\mu_1 \neq \mu_2$. Такая гипотеза соответствует двустороннему тесту. Выполняя тест, мы определяем, куда попадет t-статистика, вычисленная на экспериментальных данных. В нашем примере t-статистика = 2,2341 больше критического t_α равного 2,0244 для двустороннего теста с уровнем значимости α = 0,05.

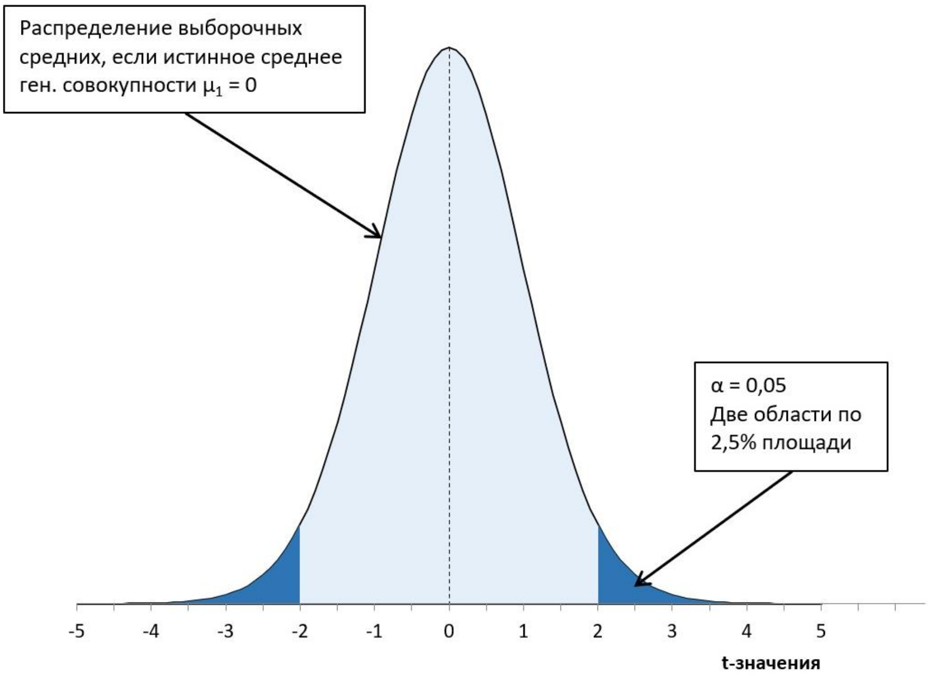


Рис. 4. Графическое представление двухвыборочного t-теста

Двухвыборочный t-тест при заданном уровне значимости α = 0,05 позволяет отвергнуть нулевую гипотезу о равенстве математических ожиданий двух генеральных совокупностей. t-тест позволяет принять альтернативную гипотезу: Выборка 1 и Выборка 2 извлечены из разных генеральных совокупностей.

Заметим, что односторонний тест используется, если в качестве альтернативной принимается одна из направленных гипотез: $\mu_1 > \mu_2$ или $\mu_1 < \mu_2$.

Ошибка II-го рода

Выше мы задали уровень значимости α = 0,05. Уровень значимости также называют ошибкой I-го рода. Значение ошибки I-го рода, равное 0,05 означает, что с вероятностью 5% мы отвергнем нулевую гипотезу, хотя она будет верна. Ошибку I-го рода нельзя сделать очень маленькой, поскольку в этом случае возрастет частота ошибок II-го рода: мы будем придерживаться нулевой гипотезы, когда она перестала быть верной. Вероятность ошибка II-го рода соответствует площади под кривой распределения выборочных средних второй группы, отсекаемых t-значением.

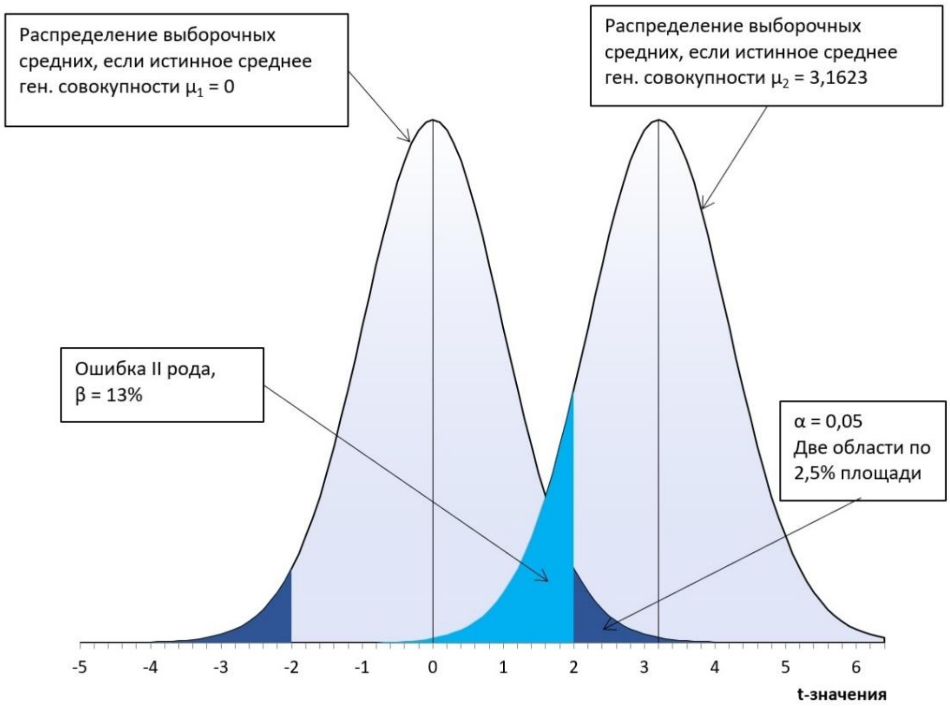


Рис. 5. Ошибка II-рода. Расчеты см. в Excel-файле.

Статистическая мощность

Статистической мощностью называется величина $w = 1 - \beta$:

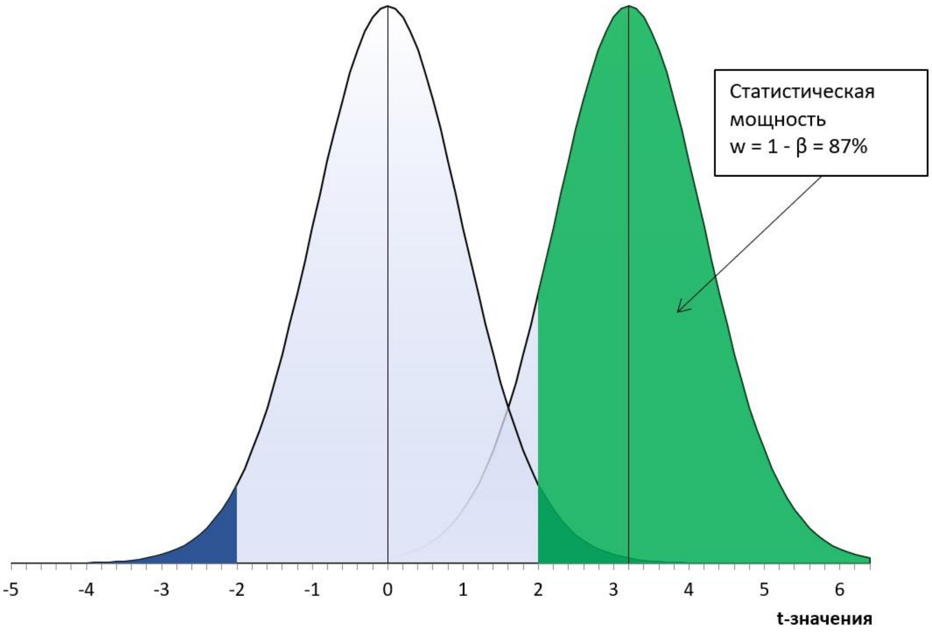


Рис. 6. Статистическая мощность

Статистическая мощность показывает, с какой вероятностью наш план эксперимента позволит отвергнуть нулевую гипотезу. Давайте проверим)) Проведем 1000 экспериментов с моделями, как на рис. 1, и подсчитаем t-значения. Можно ожидать, что около 870 t-значений будут более 2,0244. Распределение t-значений представлено на рис. 7. Этот анализ выполнен с помощью волатильных функций в Excel, так что вы можете перейти в Excel-файл на лист "Рис. 7", и понажимать клавишу F9, запуская перерасчет.

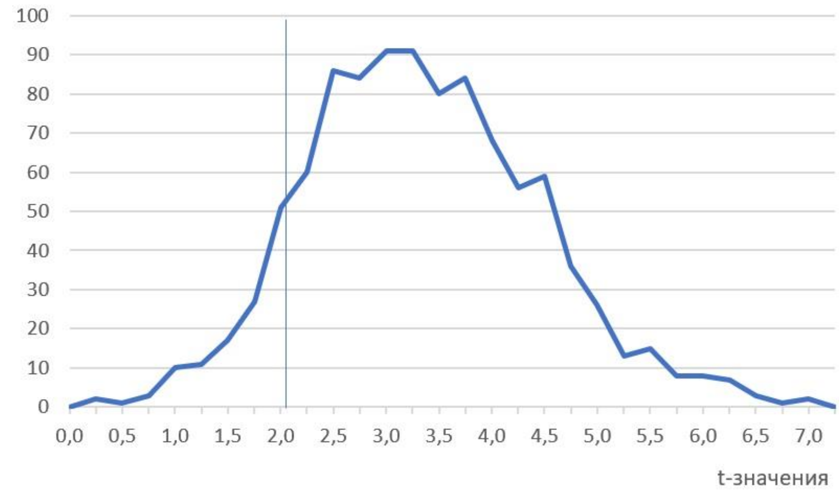


Рис. 7. Распределение t-значений в модельном эксперименте

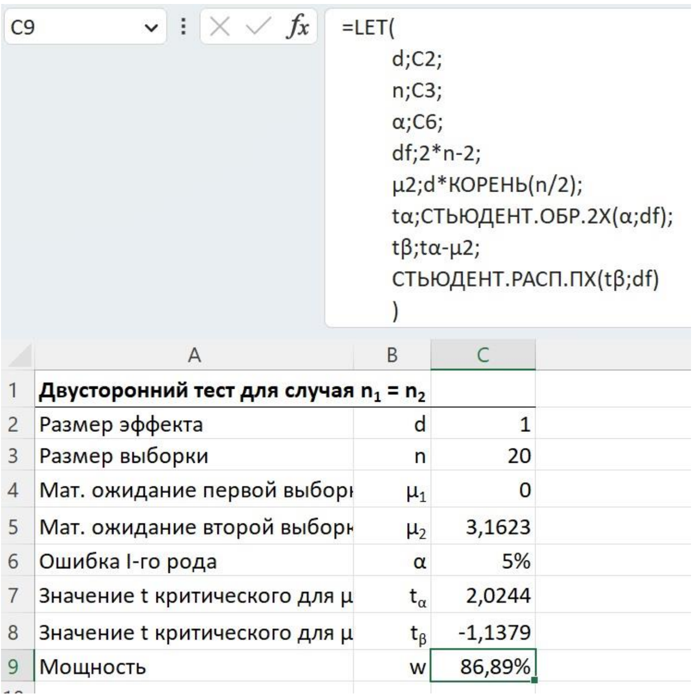
Для представленных на рис. 7 данных 874 значения из 1000 лежат в области t > 2,0244.

Аналитическое решение для вычисления статистической мощности

В общем случае (для любой альтернативной гипотезы) вычислить ошибку II-го рода и статистическую мощность не представляется возможным. Это связано с тем, что для таких вычислений требуется предположение о размере эффекта d или значении μ_2 (они взаимосвязаны). С другой стороны, оценку μ_2 дает среднее по выборке $\overline{x}_2$. Основываясь на выборочной статистике можно оценить мощность предложенного плана эксперимента.

Что это нам дает? Оценив мощность эксперимента, мы получаем вероятность, с которой можем ожидать обнаружить эффект в последующих экспериментах.

Если в общем случае рассчитать статистическую мощность не получится, то для фиксированного размера эффекта в Excel есть аналитическое решение:


C9

```
=LET(
      d;C2;
      n;C3;
      α;C6;
      df;2*n-2;
      µ2;d*КОРЕНЬ(n/2);
      tα;СТЬЮДЕНТ.ОБР.2Х(α;df);
      tβ;tα-µ2;
      СТЬЮДЕНТ.РАСП.ПХ(tβ;df)
      )
```

	A	B	C
1	**Двусторонний тест для случая $n_1 = n_2$**		
2	Размер эффекта	d	1
3	Размер выборки	n	20
4	Мат. ожидание первой выборі	μ_1	0
5	Мат. ожидание второй выборк	μ_2	3,1623
6	Ошибка I-го рода	α	5%
7	Значение t критического для µ	t_α	2,0244
8	Значение t критического для µ	t_β	-1,1379
9	Мощность	w	86,89%

Рис. 8. Формула для расчета статистической мощности в Excel

Если вы не знакомы с функцией LET, рекомендую почитать [Упрощение формул Excel путем именования фрагментов с помощью функции LET]().

Здесь функция настроена на случай, когда размеры двух сравниваемых выборок равны: $n_1 = n_2 = n$. После объявления функции LET в последующих трех строках функция считывает с листа Excel три значения: размер эффекта d, размер выборки n, вероятность ошибки I-го рода α. А далее последовательно вычисляет число степеней свободы df, матожидание распределения, соответствующего альтернативной гипотезе μ_2, критическое значение t_α, значение t_β, соответствующее t_α при переносе нуля в точку μ_2, и наконец, площадь под кривой распределения μ_2 справа от t_β. Вуаля! Это и есть статистическая мощность для заданных d, n, α.

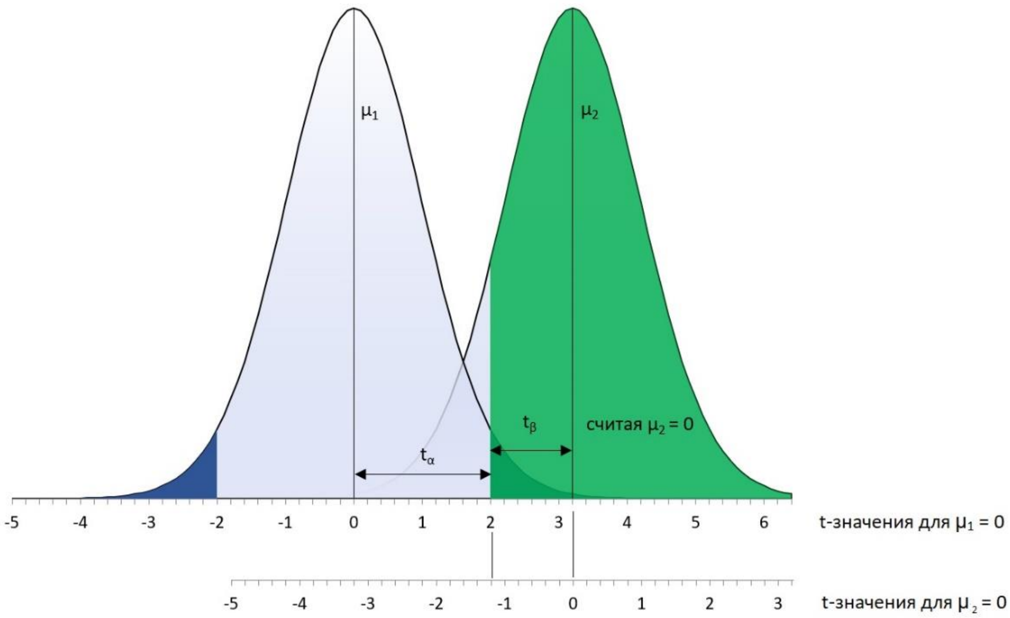


Рис. 9. Обозначения на графике статистической мощности

Расчет размера выборки при заданной статистической мощности

Для экспериментатора может быть интересна и обратная задача: для заданных d, α, w рассчитать необходимый размер выборки n. Аналитическое решение в Excel я думаю не существует. Так как есть функции, возвращающие вероятность, например СТЬЮДЕНТ.РАСП()… есть функции, возвращающие t-значение, например, СТЬЮДЕНТ.ОБР(), но… нет функций, возвращающих число степеней свободы.

Это ограничение можно обойти, построив *Таблицу данных* (меню *Данные* –> *Прогноз* –> *Анализ "что если"* –> *Таблица данных*). Таблица и график позволяют определить *n*, соответствующее требуемой мощности:

4	17,1%
6	31,5%
8	44,3%
10	55,3%
12	64,5%
14	72,0%
16	78,1%
18	83,0%
20	86,9%
22	89,9%
24	92,3%
26	94,2%
28	95,6%
30	96,7%
32	97,5%
34	98,1%
36	98,6%
38	99,0%
40	99,2%

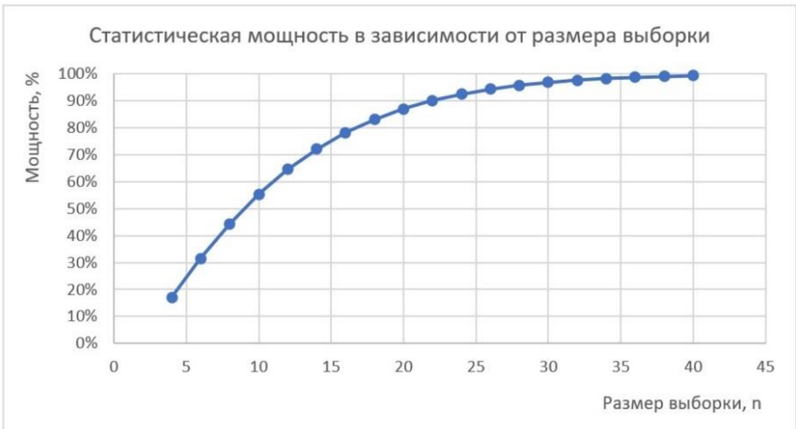


Рис. 10. Графическое решение обратной задачи: для заданных d, α, w рассчитать необходимый размер выборки n

Понятно, что четыре переменные…

- размер эффекта d
- размер выборки n
- вероятность ошибки I-го рода α

- и статистическая мощность w

… однозначно связаны. Поэтому можно сформулировать и такой вопрос: при заданных n, α, w, какой размер эффекта можно обнаружить?

Калькулятор статистической мощности

И, наконец следует упомянуть, что в Инете есть бесплатные программы – калькуляторы, рассчитывающие статистическую мощность. Например, G*Power. На рис. 11 показан итог работы этой программы. На вход были поданы следующие условия:

- двусторонний t-критерий
- размер эффекта на генеральной совокупности d = 0,55
- частота ошибок I-го рода α = 0,05
- размеры выборок $n_1 = n_2 = 40$

На выходе получены значения

- параметр нецентральности δ; он же матожидание распределения альтернативной гипотезы $\mu_2 = 2{,}4597$
- критическое t-значение, $t_\alpha = 1{,}9908$
- число степеней свободы df = 78
- статистическая мощность w = 0,68

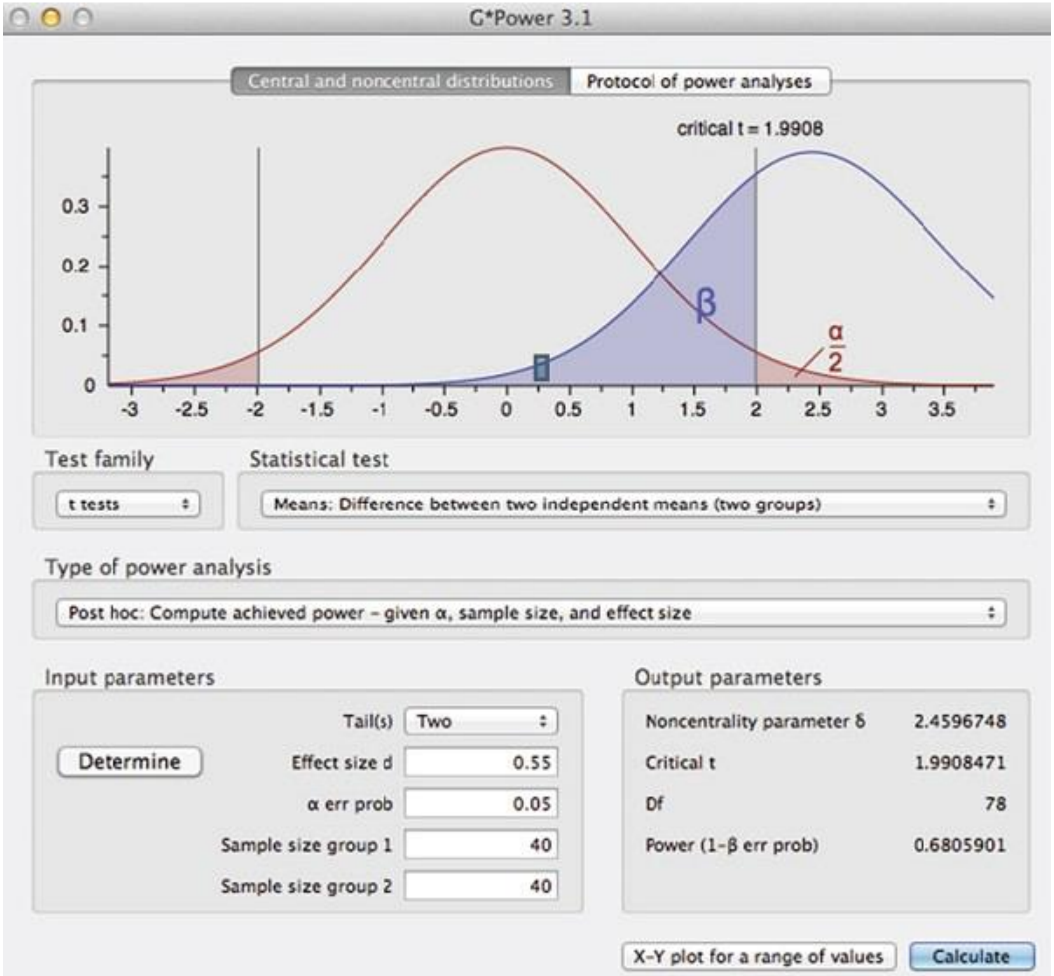


Рис. 11. Вывод программы G*Power, вычисляющей мощность t-критерия

Использованные материалы

Майкл Херцог. Статистика и планирование эксперимента для непосвященных

Statistical Power Analyses for Mac and Windows G*Power

Конрад Карлберг. Регрессионный анализ в Microsoft Excel

t-статистика Стьюдента в Excel

Упрощение формул Excel путем именования фрагментов с помощью функции LET