Рекомендации по применению тестов эквивалентности

Роберт А. Крибби

Йоркский университет

Джеймми А. Груман

Виндзорский университет

Шанталь А. Арпен-Крибби

Йоркский университет

Исследователи в области психологии надежно выбирают традиционные тесты значимости нулевой гипотезы (например, t-критерий Стьюдента), независимо от того, относится ли исследовательская гипотеза к тому, эквивалентны ли средние значения групп или различаются ли средние значения групп. Тесты эквивалентности, которые были популярны в биофармацевтических исследованиях в течение многих лет, недавно были введены и рекомендованы исследователям в области психологии для демонстрации эквивалентности двух групп средних. Однако существует очень мало рекомендаций по применению тестов эквивалентности. Исследование методом Монте-Карло использовалось для сравнения критерия эквивалентности, предложенного Шуирманном, с традиционным t-критерием Стьюдента для определения эквивалентности двух групповых средних. Было обнаружено, что критерий эквивалентности Шуирмана более эффективен, чем t-критерий Стьюдента, для определения эквивалентности средних значений генеральной совокупности при больших размерах выборки; однако критерий эквивалентности Шуирмана плохо работает по сравнению с t-критерием Стьюдента при небольших размерах выборки и/или завышенных дисперсиях. © 2003 Wiley Periodicals, Inc. J Clin Psychol 60: 1–10, 2004.

Ключевые слова: проверка нулевой гипотезы; тесты эквивалентности; дисперсионная неоднородность; Студенческая т

Исследователи в области психологии, которые заинтересованы в сравнении средних значений двух групп по зависимому показателю, надежно выбирают традиционные тесты значимости нулевой гипотезы (например, t-критерий Стьюдента), в которых нулевая гипотеза относится к эквивалентности населения.

Корреспонденцию по поводу этой статьи следует направлять по адресу: Robert A. Cribbie, факультет психологии Йоркского университета, 4700 Keele Street, Toronto, ON M3J 1P3, Canada; электронная почта: cribbie@yorku.ca.

Опубликовано в Интернете в Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jclp.10217

означает. Кроме того, традиционные тесты нулевой гипотезы надежно применялись независимо от
относится ли исследовательская гипотеза к тому, отличаются ли групповые средние значения или
эквивалент. Например, клинический исследователь может быть заинтересован в оценке исследования.
гипотеза о том, что анорексики, страдающие анорексией, страдающие перееданием/очищением, и анорексики, ограничивающие себя, имеют одинаковые уровни
устойчивость к лечению. Эта исследовательская гипотеза вне отличается от гипотезы о том, что анорексики,
страдающие анорексией от переедания/очищения, и анорексики, ограничивающие анорексию, имеют разные уровни
устойчивости к лечению и, как будет показано ниже, могут иметь важные последствия для статистических данных.
приняла процедура.

Андерсон и Хаук (1983), Руанет (1996), Шуирманн (1987), Селвин и Холл
(1984), Westlake (1976) и другие предложили статистические методы для определения того,
две группы эквивалентны по конкретной зависимой мере, где исследователь априори определяет минимальное
различие, приемлемое для объявления групповых средних эквивалентными. Эти методы недавно были представлены
исследователям в области психологии через
влиятельные статьи Роджерса, Ховарда и Весси (1993) и Симэна и Серлина (1998).
Обе статьи посвящены критерию эквивалентности, предложенному Шуирманном (хотя, как
Seaman & Serlin отмечают, Rogers et al. отдайте должное Westlake, 1976 за этот метод).
Тест эквивалентности Шуирмана был чрезвычайно популярен в биофармацевтике.
исследования для демонстрации биоэквивалентности, хотя до недавнего времени тесты эквивалентности
редко принимались исследователями в области психологии, хотя исследовательские гипотезы, касающиеся
эквивалентности, часто исследуются (Rogers et al., 1993).

Есть по крайней мере две основные причины рекомендовать тесты эквивалентности.
Во-первых, цель исследования может заключаться не в том, чтобы показать, что методы лечения идентичны, а только в том, чтобы показать, что
различия между методами лечения слишком малы, чтобы их можно было считать значимыми. Рассмотрим, например,
клинического психолога, заинтересованного в изучении двух конкурирующих методов лечения.
для депрессии - один для долгосрочной терапии и один для краткосрочной терапии. Исследователь
может быть интересно продемонстрировать, что результаты лечения при краткосрочной терапии эквивалентны
результатам долгосрочной терапии, в дополнение к тому, что она занимает меньше времени, дешевле и т. д. В этом
примере исследователю может не понадобиться показывать
что методы лечения «абсолютно эквивалентны» (как и в случае с традиционной нулевой гипотезой, Но:
м1 м2 ! но только то, что различия в результатах лечения недостаточно велики, чтобы гарантировать
принятие более трудоемкой и дорогой терапии (т. е. @m1 m2 # D, где
D представляет собой априорную критическую разницу для определения эквивалентности). Во-вторых, хорошо
Известно, что по мере увеличения размера выборки вероятность нахождения даже мельчайших (и потенциально
бессмысленных) средних различий статистически значимыми приближается к единице с традиционными
тестами нулевой гипотезы эквивалентности, такими как две независимые выборки Стьюдента
тест. Это особенно важно, учитывая возрошее количество запросов, например, от учебников, редакторов журналов и
статистиков, чтобы исследователи обосновали значимость своих результатов, включив такие меры величины
эффекта, как d, h2 и т. д. (см.
Боуг и Томпсон, 2001; Томпсон, 2002a, 2002b). Поэтому при большой выборке
Исследователи размеров могут захотеть указать критическую разницу между методами лечения, которые могли бы
считать клинически значимым.

Роджерс и др. (1993) продемонстрировали, как результаты исследований эквивалентности двух
экспериментальных групп с использованием методов проверки нулевой гипотезы неэквивалентности часто
противоречат результатам, полученным с помощью традиционной нулевой гипотезы эквивалентности.
тесты. Например, Роджерс и др. сравнивали испытуемых, пристрастившихся к алкоголю, и испытуемых
пристрастие к наркотикам по шкале коррекции MMPI (на основе исследования Кэннона, Белла,
Fowler, Penk, & Finkelstein, 1990), используя как критерий эквивалентности, предложенный Шуирманом, так и t-
критерий Стьюдента. Авторы обнаружили, что две группы средних были объявлены
статистически равные по t-критерию Стьюдента, но статистически эквивалентные по критерию Шуирмана

проверка эквивалентности. Авторы также обнаружили, что по шкале шизофрении MMPI
две группы оказались статистически неразличимыми по t-критерию Стьюдента, но не эквивалентными по критерию
эквивалентности Шуирмана.

Учитывая недавние рекомендации в поддержку проверки эквивалентности и увеличение
наличие тестов эквивалентности в психологических исследованиях, важно, чтобы исследователи
иметь четкие руководящие принципы для применения тестов эквивалентности, а также когда тесты эквивалентности
могут быть неуместными. Например, на иллюстрации, представленной ранее, нет
способ узнать, эквивалентны ли баллы по шкале MMPI у субъектов, пристрастившихся к алкоголю, и у субъектов,
зависимых от наркотиков, и, таким образом, нет никакого способа узнать, являются ли результаты
Критерий эквивалентности Шуирмана верен или если результаты традиционного теста Стьюдента t
тест верный. Поэтому целью настоящего исследования является сравнение рекомендуемого в настоящее
время метода оценки эквивалентности двух групповых средних, предложенного Шуирманном (1987), с традиционным
методом t-критерия Стьюдента. Следующее обсуждение будет
(a) рассмотреть критерий эквивалентности Шуирмана, (b) обсудить применение критерия Шуирмана
тест эквивалентности в примере, предоставленном Симаном и Серлином (1998, стр. 405), и
(c) использовать исследование Монте-Карло для сравнения метода Шуирманна с t-критерием Стьюдента.
метод оценки средней эквивалентности генеральной совокупности.

Критерий эквивалентности Шуирмана

Первым шагом в проведении теста эквивалентности Шуирмана является установление критического среднего значения.
разница для объявления двух популяций означает эквивалентность ~ D!. Любая средняя разница меньше
чем D считалось бы бессмысленным в рамках эксперимента. Выбор интервала эквивалентности ~D! является
важным аспектом проверки эквивалентности, который
в первую очередь зависит от субъективного «уровня уверенности», с которым можно заявить о двух (или более)
эквивалент населения. Этот уровень достоверности может принимать множество различных форм, включая
необработанное значение (например, средние результаты теста, отличные от десятибалльных), процентную разницу (например,
610%), процент разницы объединенного стандартного отклонения и так далее. Трион (2001)
описал этот уровень достоверности как «сумму, которая считается несущественной» (стр. 379),
и Роджерс и др. (1993) заявили, что «любая разница, достаточно малая, чтобы попасть в этот интервал эквивалентности,
будет считаться клинически и/или практически неважной» (стр. 553).
рекомендуют исследователям, обсуждающим подходящее значение D, учитывать природу
исследовательская работа. Например, если долговременная терапия, о которой говорилось ранее, занимала в три раза больше времени,
и была в три раза дороже, чем краткосрочная терапия, то разница более существенная
в исходах (например, 20%) может потребоваться, чтобы сделать вывод, что методы лечения эквивалентны
если бы длительная терапия занимала в полтора раза больше времени и была примерно эквивалентна по стоимости
к краткосрочной терапии (где увеличение исходов на 5% может быть целесообразным для
сделать вывод, что методы лечения эквивалентны).

Предполагается, что две выборки случайным образом и независимо выбираются из нормально распределенных
совокупностей с одинаковой дисперсией. Можно провести две односторонние проверки гипотез.
используется для установления эквивалентности, где нулевая гипотеза относится к неэквивалентности
население означает и может быть выражено в виде двух отдельных составных гипотез:

$$Ho1: м1 м2 Д; Ho2: м1 м2 \quad Д+$$

Отказ от Ho1 означает, что m1 m2 D, а отказ от Ho2 означает, что m1 m2
D. Далее, отказ от обеих гипотез означает, что m1 m2 попадает в границы
из ~ Д, Д! и средства считаются эквивалентными.

Но1 отклоняется, если t1 тн где:

$$t1 = \frac{-\text{Опыт } 1 \quad X\ P\ 2\ !\ Д}{\sqrt{\dfrac{\sim n1\ n2!@\sim n1\ 1!s1^{2} \quad \sim n2\ 1!s2^{2}}{n1n2\sim n1\ n2\ 2!}}},$$

и Но2 отвергается, если t2 tn где:

$$t2 = \frac{-\text{Опыт } 1 \quad X\ P\ 2\ ! \quad \sim Д!}{\sqrt{\dfrac{\sim n1\ n2!@\sim n1\ 1!s1^{2} \quad \sim n2\ 1!s2^{2}}{n1n2\sim n1\ n2\ 2!}}}$$

1 и XP 2 — групповые средние значения, n1 и n2 — объемы групповой выборки, s1 и s2 — Стандартные отклонения группы XP и tna — критическое значение t вверх него хвостового уровня с n n1 n2 2 степенями свободы.

Моряки и Серлин (1998) Пример

Критерий эквивалентности Шуирмана был продемонстрирован Симаном и Серлином (1998, стр. 405). в следующем примере. Критическая разница ~D! для объявления населения средний эквивалент был установлен на уровне 5, номинальный уровень значимости был установлен на уровне 0,05 и:

$$XP_1 \quad 65{+}7,\ n1\ 25,\ n1\ 9$$

$$XP_C \quad 65{+}0,\ нс\ 25,с\ 2\ 8{+}$$

Подстановка выборочной статистики генерирует следующую тестовую статистику:

$$t1 \quad \frac{\sim 65{+}7\ 65{+}0!\ 5}{2{+}4} \quad 1{+}8$$

$$t2 \quad \frac{\sim 65{+}7\ 65{+}0!\ \sim 5!}{2{+}4} \quad 2{+}4{+}$$

Так как t1 ~ 1+8! а (1,68) и t2 ~2+4! тн а (1.68) средства населения объявляются эквивалентны (т. е. разница между средними не превышает критической разницы).

Симэн и Серлин также объясняют, что для этого примера такое же решение было бы было достигнуто, если применялся традиционный t-критерий Стьюдента с нулевой эквивалентностью гипотеза. В частности, нулевая гипотеза ~Но: m1 m2 ! не был бы отвергнут в этом примере, и средства снова были бы объявлены статистически эквивалентными.

Интересный вопрос, вытекающий из предыдущего примера, заключается в том, как часто два методы (критерий эквивалентности Шуирмана и t-критерий Стьюдента) приводят к одним и тем же выводам (или, что более важно, к тому, как часто методы приводят к разным выводам). Этот вопрос имеет первостепенное значение, учитывая, что исследователи в области психологии регулярно применять t-критерий Стьюдента, даже если исследовательская гипотеза касается эквивалентности означает. Имитационное исследование использовалось для определения вероятности объявления средних значений населения эквивалентными критерию эквивалентности Шуирмана (т. е. Но отвергается) и Критерий Стьюдента (т. е. Но не отвергается), где размеры выборки и статистика из предыдущий пример использовался в качестве параметров совокупности в моделировании. Это ожидаемо что критерий эквивалентности Шуирмана объявит, что средства населения эквивалентны более чаще, чем критерий Стьюдента, учитывая, что альтернативная гипотеза критерия Шуирмана ~На: мл мСм 5) охватывает большую область, чем нулевая гипотеза для Стьюдента

Таблица 1

Имитационное исследование вероятности обнаружения эквивалентности совокупности с использованием параметров (m1 65, m2 65,7, s1 8, s2 9, D 5) Из Seaman and Serlin (1997, стр. 405)

	Размер выборки			
Статистический тест	n1 n2 25a	n1 n2 50	n1 n2 75	n1 n2 100
Критерий эквивалентности Шуирмана	.311	0,768	0,913	0,974
t - критерий Стьюдента	0,939	0,934	0,914	0,912

а Фактические размеры выборки, использованные в примере Симена и Серлина.

t test ~Ho: m1 m2 0), а также потому, что разница между средними значениями (65,7 65,0 0,7) я вно нах одится в пределах критической разницы D 5. Было выполнено пять тысяч симуляций. проводится с использованием номинального уровня значимости 0,05. Результаты моделирования представлены в таблице 1. Вопреки логике, лежащей в основе двух методов, результаты показывают, что со средними значениями, дисперсиями и размерами выборки Симена и Серлина Например, вероятность объявить средства эквивалентными t Стьюдента (0,939) была значительно больше, чем с критерием эквивалентности Шуирмана (0,311). Далее, даже когда размеры выборки были увеличены до n 50 и n 75, критерий Стьюдента был по крайней мере так же вероятно, как критерий эквивалентности Шуирмана, чтобы объявить средства эквивалентными. Нет, это не так пока размеры выборки не были увеличены до n 100, критерий эквивалентности Шуирмана стал более вероятным, чем t - критерий Стьюдента, чтобы объявить средства эквивалентными. Однако эти результаты основаны на конкретных параметрах населения и могут не отражать общую эффективность процедур. Поэтому требуется более подробное сравнение критерия эквивалентности Шуирмана и t - критерия Стьюдента.

Исследование Монте-Карло

Имитационное исследование использовалось для сравнения критерия эквивалентности Шуирманна с традиционным t - критерием Стьюдента для обнаружения эквивалентности населения в условиях, обычно опыт исследователей в области психологии. При этом манипулировались несколькими переменными. исследование, включая (а) размер выборки, (б) среднюю конфигурацию населения и (в) население отклонения. Критическая средняя разница для установления эквивалентности популяции с помощью критерия эквивалентности Шуирмана поддерживалась на уровне 1 во всех условиях.

Одной из основных причин использования тестов эквивалентности является то, что в качестве выборки размер увеличивается, вероятность того, что даже тривиальные средние различия будут статистически значимыми, становится большой. Таким образом, размеры выборки группления были изменены в этом исследовании. исследование. В частности, размеры групповой выборки были установлены на уровне n1 n2 10, n1 n2 25, n1 n2 50 и n1 n2 100.

Эффективность проверки эквивалентности напрямую зависит от неоднородности значит группа. В этом исследовании оценивались пять средних конфигураций, включая эквивалентные средние значения совокупности ~ m1 m2 ! а четыре неэквивалентные совокупности означают ~ m1 m2 +4, m1 m2 +8, m1 m2 1+2 и m1 m2 1,6). Учитывая, что критическая разница если эквивалентность совокупности установлена равной 1, конфигурация эквивалентных средних и неэквивалентные конфигурации с m2 m1 1 подпадают под альтернативную гипотезу критерия эквивалентности Шуирмана (т. е. разность средних значений совокупности не превышает критическая средняя разница, и, таким образом, ожидается, что средние значения будут объявлены эквивалентными

критерий Шуирмана), а ненулевые конфигурации с m2 m1 1 попадают под нулевой гипотезой критерия эквивалентности Шуирмана (т. е. разность средних значений генеральной совокупности превышает разность критических средних, и, таким образом, ожидается, что средние значения будут объявлены неэквивалентными критерию Шуирмана). Для t-критерия Стьюдента разница средних значений в любой популяции больше чем ноль подпадает под альтернативную гипотезу ~ m1 m2 ! и предполагается, что средства объявлен неэквивалентным.

Другой важный вопрос заключается в том, какой эффект будет иметь увеличение изменчивости населения? имеют по критерию эквивалентности Шуирмана. Также в отношении дисперсии населения, в обзоре опубликованных исследований в области образования и психологии, Keselman et al. (1998) найдено что неравные отклонения были нормой, а не исключением. В частности, исследователи часто сообщают об отношении наибольшей дисперсии к наименьшей величине, равной четырем, и наибольшей к наименьшей коэффициенты дисперсии, достигающие восьми, не были редкостью. Таким образом, в дополнение к исследованию случая, когда дисперсия обеих групп была установлена равной единице, влияние популяции инфляция дисперсии также была исследована в этом исследовании путем установки дисперсии одного из группы до четырех-восьми.

Для каждого условия было проведено пять тысяч симуляций с использованием номинального уровня значимости 0,05.

Полученные результаты

Априорная эквивалентность

В табл.

априорные средние различия в популяции были меньше, чем критическая средняя разница, выборка размер был основным фактором при сравнении критерия эквивалентности Шуирманна и критерия Стьюдента. критерий (Здесь следует отметить, что для t-критерия Стьюдента любая популяционная разность средних

Таблица 2

Вероятность обнаружения популяционной эквивалентности для критерия эквивалентности Шуирмана (критическая разница для эквивалентности 1) и критерий Стьюдента

		2 с 1 2 с 2 1		2 с 1 0c 2 2 4		2 с 1 0c 2 2 8	
н	м1 м2	С	т	С	т	С	т
10	0	.3930	.9488	.0248	.9466	.0026	.9412
	0,4	.2764	.8576	.0220	.9076	.0016	.9234
	0,8	.0966	.5962	.0114	.7940	.0008	0,8582
25	0	.9384	.9534	.4426	.9506	.0698	0,9488
	0,4	.6778	.7174	.3016	.8314	.0526	0,8922
	0,8	.1698	.2036	.1056	.5832	.0308	.7528
50	0	.9996	.9510	.8660	.9494	.5230	.9514
	0,4	.9100	.4962	.5922	.7676	.3512	.8530
	0,8	.2562	.0202	.1544	.2974	.1174	.5320
100	0	1.000	.9496	.9946	.9510	.9083	0,9498
	0,4	.9956	.1922	.8444	.5688	.6324	.7330
	0,8	.3998	.0000	.2388	.0544	.1714	.2526

Критерий эквивалентности С. Шуирмана; t - критерий Стьюдента.

делает нулевую гипотезу ложной, и поэтому цель проверки при всех ненулевых условиях состоит в том, чтобы выявить неэквивалентность, а не эквивалентность средних.

Доля неотвержений нулевой гипотезы для t Стьюдента с средними различиями больше нуля [т. е. ошибки типа II] представлены для сравнения только с процедурой Шуирмана.) При 10 или 25 наблюдениях в группе критерий Стьюдента был более вероятным. объявить две группы означает эквивалентность, чем критерий эквивалентности Шуирмана, независимо от размера разности средних значений генеральной совокупности. Фактически, с 10 субъектами в группе тест эквивалентности Шуирмана никогда не обнаруживал эквивалентности более чем в 50% случаев, тогда как тест Стьюдента никогда не обнаруживал эквивалентности менее чем в 50% случаев.

С другой стороны, при 50 или 100 субъектах в группе критерий эквивалентности Шуирманна с большей вероятностью, чем критерий Стьюдента, выявлял эквивалентность.

Когда дисперсии генеральной совокупности были завышены в одной группе, критерий Шуирманна очень плохо определял эквивалентность средних значений генеральной совокупности. В частности, когда дисперсия одной группы была установлена равной восьми, критерий эквивалентности Шуирманна никогда не был более эффективным для обнаружения эквивалентности, чем t-критерий Стьюдента, независимо от размера выборки или конфигурации средней совокупности. Когда дисперсия одной группы была установлена равной четырем, критерий эквивалентности Шуирманна был более эффективным, чем t-критерий Стьюдента, только когда в группе было не менее 100 субъектов. Когда средние группы населения были равны, критерий Стьюдента, как и ожидалось, обнаружил, что средние группы эквивалентны примерно в 95% случаев, независимо от размера выборки. С другой стороны, критерий эквивалентности Шуирманна часто обнаруживал, что средние значения группы неэквивалентны, даже когда средние значения совокупности были идентичными. Например, при 10 субъектах в группе, равных средних значениях x популяции и дисперсии одной группы, равной четырем, тест эквивалентности Шуирманна показал, что средние значения эквивалентны только в 2,48% случаев, тогда как критерий Стьюдента обнаружил, что средние значения эквивалентны в 2,48% случаев. 94,66% случаев. Увеличение размера выборки улучшило производительность теста Шуирмана, хотя тест по-прежнему работал плохо по сравнению с t Стьюдента. Например, при 50 субъектах в группе, равных средних значениях x популяции и дисперсии одной группы, равной четырем, тест эквивалентности Шуирманна показал, что средние значения эквивалентны в 86,60% случаев, тогда как t Стьюдента обнаружил, что средние значения эквивалентны в 94,84. % случаев.

Априорная неэквивалентность

Вероятности обнаружения неэквивалентности популяции с помощью критерия эквивалентности Шуирмана и t-критерия Стьюдента для смоделированных условий представлены в табл. 3. все случаи) при обнаружении различий.

Однако ясно, что превосходная способность критерия Шуирманна обнаруживать средние различия, превышающие критическую разницу, является функцией систематической ошибки теста для объявления популяций неэквивалентными, даже когда различия были меньше критической разницы. Мощность t Стьюдента для обнаружения средних различий, как и ожидалось, зависит от размера выборки и дисперсии, при этом мощность максимизируется при больших размерах выборки и небольших дисперсиях.

Обсуждение

В настоящей статье исследована альтернатива традиционному t-критерию Стьюдента для определения эквивалентности средних значений двух лечебных групп. Есть много примеров (представленных здесь и в других местах) парадигм исследования клинической психологии, в которых интерес представляет вопрос, является ли одно лечебное средство практически эквивалентным второму среднему, или в другом

Таблица 3

Вероятность обнаружения неэквивалентности популяции для критерия эквивалентности Шуирмана (критическая разница для эквивалентности 1) и критерий Стьюдента

		$2c1 \quad 2c2 \quad 1$		$2c1^{0c}2^{2} \quad 4$		$2c1^{0c}2^{2} \quad 8$	
н	м1 м2	С	т	С	т	С	т
10	1,2	.9832	.7254	.9928	.3598	.9982	.2290
	1,6	.9980	.9280	.9974	.5890	.9992	.3780
25	1,2	.9900	.9874	.9834	.7486	.9886	.5006
	1,6	.9998	1.000	.9988	.9414	.9980	.7446
50	1,2	.9942	.9998	.9856	.9584	.9798	0,7928
	1,6	1.000	1.000	.9998	.9990	.9990	0,9588
100	1,2	.9984	1.000	.9924	.9994	.9876	0,9764
	1,6	1.000	1.000	1.000	1.000	1.000	.9998

Критерий эквивалентности С. Шуирмана; t - критерий Стьюдента.

Другими словами, разница между двумя лечебными средствами не настолько велика, чтобы считаться осмысленный. В недавних статья х исследователя м в области психологии были представлены тесты эквивалентности и рекомендованы эти процедуры для ответов на вопросы, касающиеся эквивалентности двух групповых средних. Эти статьи увеличили как доступность, так и популярность этих процедур. Однако статистических исследований мало. практически отсутствуют свойства процедур и руководства по применению тестов эквивалентности.

В этой статье были проведены два исследования моделирования: (a) расширение примера, использованного Симаном и Серлином (1998) для демонстрации применения теории Шуирманна. критерий эквивалентности и (b) сравнение критерия эквивалентности Шуирманна и t - критерия Стьюдента во многих условия х, обычно встречающих ся в экспериментах поповеденческим наукам. Цель исследований моделирования состояла в том, чтобы выделить статистические свойства тесты эквивалентности и как они соотносятся с традиционной проверкой нулевой гипотезы. Так должно быть отметил, что многие из результатов, представленных в этой статье, можно было бы предсказать на основе лежащих в основе выборочных распределений тестовой статистики, хотя важно, чтобы эти результаты должны быть количественно оценены, чтобы позволить исследователя м принимать обоснованные решения при выборе соответствующую тестовую статистику. Оба исследования показали, что размер выборки я вляется решающим фактором. при выборе между критерием эквивалентности Шуирмана и традиционным t - критерием Стьюдента. Если число субъектов на одно условие велико (25 или более), критерий эквивалентности Шуирманна может быть более подходящим для обнаружения эквивалентности населения, чем критерий Стьюдента, особенно, когда различия средних по популяции присутствуют, но меньше критической разницы. По мере увеличения размера выборки и разницы между средними t критерий Стьюдента как и ожидалось, становится более мощным в обнаружении различий и, следовательно, с меньшей вероятностью объявить различия бессмысленными. С другой стороны, по мере увеличения размера выборки критерий эквивалентности Шуирманна становится более мощным при обнаружении этих различий. меньше, чем критическая разница, и поэтому рекомендуется для t Стьюдента с большими размеры выборки.

Однако требуется важная оговорка к предыдущей рекомендации относительно размера выборки; когда групповые дисперсии даже умеренно завышены, способность Процедура Шуирмана для обнаружения эквивалентности существенно сокращена. Хотя это открытие частично является результатом снижения мощности (т. е. увеличения стандартной ошибки)

Стьюдента для обнаружения различий средних с завышенными дисперсиями, не следует упускать из виду экстремальный эффект завышения дисперсий на критерий эквивалентности Шуирмана (особенно при небольших размерах выборки). В частности, крайне проблематично то, что при отсутствии различий между средними значениями (и вероятность вывода об эквивалентности группс t Стьюдента составляет приблизительно 0,95) критерий эквивалентности Шуирмана так плохо работает с малыми размерами выборок и/или увеличение отклонений.
Кроме того, также важно признать, что, хотя это и не является предметом этой статьи, неравные размеры выборки в сочетании с неравными дисперсиями могут значительно повлиять на частоту ошибок типа I и типа II для t Стьюдента, помимо того, что наблюдалось в этом исследовании. исследование. Ожидается, что влияние неравных размеров выборок и дисперсий также значительно повлияет на частоту ошибок типа I и типа II критерия эквивалентности Шуирмана, и поэтому необходимо исследование надежного теста эквивалентности для неравных размеров выборки и дисперсии.

Хотя результаты этого исследования важны с точки зрения предоставления рекомендаций исследователя относительно выбора подходящей тестовой статистики для оценки эквивалентности двух групповых средних, существуют два важных ограничения текущего исследования, которые следует учитывать. Во-первых, исследование Монте-Карло не смогло изучить все возможные условия размера выборки, дисперсионного неравенства и т. д.; поэтому, хотя мы ожидаем, что результаты этого исследования будут обобщены на многие распространенные среды тестирования, эти результаты специфичны для условий, изучаемых в этом исследовании. Во-вторых, хотя в этой статье мы сосредоточились на схеме проверки гипотез для оценки эквивалентности средних, были достигнуты важные успехи в применении подходов доверительного интервала к проверке эквивалентности, которые здесь не рассматривались, но могут представлять интерес. читателям (см., например, Seaman & Serlin, 1998; Tryon, 2001).

Подводя итог, можно сказать, что тесты на эквивалентность чрезвычайно популярны в биофармацевтических исследованиях для демонстрации того, что эффекты двух лекарств практически эквивалентны. Ожидается, что по мере роста числа исследований, описывающих методологию тестов эквивалентности, популярность тестов эквивалентности в области психологии будет расти, учитывая, что исследователи будут более подготовлены к выявлению ситуаций, в которых тесты эквивалентности уместны. Поэтому важно, чтобы существовали четкие рекомендации по применению этих тестов. Результаты этого исследования подчеркивают необходимость признания того, что критерий эквивалентности Шуирмана и критерий Стьюдента диаметрально противоположны в своем подходе к проверке гипотез, и, следовательно, одни и те же факторы, которые существенно влияют на способность критерия Стьюдента t обнаруживать различия между средними значениями (например, , размер выборки, вариабельность ошибок) также существенно влияют на способность теста Шуирманна обнаруживать эквивалентность.

использованная литература

Андерсон, С.А., и Хаук, В.В. (1983). Новая процедура проверки эквивалентности в сравнительных исследованиях биодоступности и других клинических исследованиях. Коммуникации в статистике: теория и методы, 12, 2663–2692.

Боуг, Ф., и Томпсон, Б. (2001). Использование размеров эффекта в исследованиях в области социальных наук: новые требования APA и журналов для улучшения методов методологии. Журнал исследований в области образования, 11, 120–129.

Кэннон, Д.С., Белл, В.Е., Фаулер, Д.Р., Пенк, В.Е., и Финкельштейн, А.С. (1990). Различия MMPI между алкоголиками и наркоманами: влияние возраста и расы. Психологическая оценка: журнал клинической и консультационной психологии, 2, 51–55.

Кесельман, Х. Дж., Хьюберти, С. Дж., Ликс, Л. М., Олейник, С., Кribби, Р. А., Донахью, Б., Ковальчук, Р. К., Лоуман, Л. Л., Петоски, М. Д., Кесельман, Дж. К., и Левин, младший (1998). Статистические практики исследователей в области образования: анализ их ANOVA, MANOVA и ANCOVA анализов. Обзор образовательных исследований, 68, 350–386.

Роджерс, Дж. Л., Ховард, К. И. и Весси, Дж. Т. (1993). Использование критериев значимости для оценки эквивалентности между двумя экспериментальными группами. Психологический бюллетень, 113, 553–565.

Руане, Х. (1996). Байесовские методы оценки важности эффектов. Психологический вестник, 119, 149–158.

Шуирманн, ди-джей (1987). Сравнение процедуры двухсторонних тестов и степенного подхода для оценки эквивалентности средней биодоступности. Журнал фармакокинетики и биофармацевтики, 15, 657–680.

Моряк, Массачусетс, и Серлин, Р.К. (1998). Доверительные интервалы эквивалентности для двухгрупповых сравнений сыновья средства. Психологические методы, 3, 403–411.

Селвин, М.Р., и Холл, Н.Р. (1984). Обайесовских методах биоэквивалентности. Биометрия, 40, 1103–1108 гг.

Томпсон, Б. (2002a). Какими могут быть будущие количественные исследования в области социальных наук: доверительные интервалы для величины эффекта. Исследователь в области образования, 31, 24–31.

Томпсон, Б. (2002b). «Статистические», «практические» и «клинические»: сколько видов значимости консультанты должны учитывать? Журнал консультирования и развития, 80, 64–71.

Трион, WW (2001). Оценка статистической разницы, эквивалентности и неопределенности с использованием доверительных интервалов вывода: интегрированный альтернативный метод выдвижения нулевой гипотезы статистические тесты. Психологические методы, 6, 371–386.

Вестлейк, WJ (1976). Симметричные доверительные интервалы для исследований биоэквивалентности. Биометрика, 37, 589–594.