Routledge
Taylor & Francis Group

Журнал экспериментального образования

ISSN: 0022-0973 (печать) 1940-0683 (онлайн) Домашняя страница журнала: http://www.tandfonline.com/loi/vjxe20

Эквивалентность дисперсий населения: Синхронизация цели и анализа

Констанс А. Мара и Роберт А. Крибби

ЖУРНАЛ ЭКСПЕРИМЕНТАЛЬНОГО ОБРАЗОВАНИЯ
http://dx.doi.org/10.1080/00220973.2017.1301356

ИЗМЕРЕНИЯ, СТАТИСТИКА И ДИЗАЙН ИССЛЕДОВАНИЯ

Эквивалентность дисперсии населения: синхронизация цели и анализа

Констанс А. Мараа и Роберт А. Криббиб

[а]Поведенческая медицина и клиническая психология, Медицинский центр детской больницы Цинциннати, Огайо, США; [б]Количественный
Программа методов, факультет психологии, Йоркский университет, Торонто, Онтарио, Канада

АННОТАЦИЯ
Исследователи часто заинтересованы в установлении эквивалентности дисперсий населения. Традиционные процедуры, основанные на различиях, подходят для ответа на вопросы о различиях в некоторых статистических показателях (например, отклонениях и т. д.). Однако, если исследователь заинтересован в оценке эквивалентности дисперсий генеральной совокупности, целесообразнее использовать процедуру, предназначенную для определения эквивалентности. Имитационное исследование использовалось для сравнения новых тестов на основе эквивалентности с традиционными тестами на однородность дисперсии при общих условиях данных. Результаты показали, что традиционные тесты, основанные на различиях, оценивают равенство дисперсий с неправильной точки зрения и что предложенный тест Левена-Веллека-Уэлча на эквивалентность групповых дисперсий является наиболее эффективным тестом для обнаружения эквивалентности. Функция R предназначена для облегчения использования этого теста на эквивалентность дисперсий населения.

КЛЮЧЕВЫЕ СЛОВА
дисперсионный анализ; дисперсия; проверка эквивалентности; однородность дисперсий

ОДНОРОДНОСТЬ ДИСПЕРСИЙ возникает, когда распределения населения имеют одинаковую дисперсию. Исследователи все больше интересуются свойствами своих данных помимо центральной тенденции, такой как дисперсия. Например, Borkenau, Hrebıckov, Kuppens, Realo и Allik (2013) выдвинули гипотезу о том, что самооценка личности будет иметь одинаковую вариабельность у мужчин и женщин. Salgado (1995) исследовал, была ли вариабельность коэффициентов валидности в тестах самоотчета для конкретной конструкции эквивалентна вариабельности коэффициентов валидности в психомоторных тестах, оцененных внешним оценщиком той же конструкции. Более известная причина изучения дисперсий — проверка предположения об однородности дисперсий, связанного с традиционными параметрическими тестами разностей средних. Независимо от причины, исследователям нужен действующий тест для оценки вопросов, связанных с изменчивостью.

Были проведены обширные исследования различных тестов, которые можно использовать для проверки различий в дисперсиях. В этой статье обсуждается, решают ли традиционные тесты однородности дисперсии проблему равенства дисперсий с неправильной точки зрения. Мы утверждаем, что для проверки однородности дисперсии следует использовать тесты эквивалентности, потому что исследовательская гипотеза равенства дисперсий правильно согласуется с альтернативной гипотезой, а не с нулевой гипотезой. С этой целью мы сначала поместим тест на равенство групповых дисперсий в рамках проверки эквивалентности. Несмотря на то, что процедуры, основанные на различиях, подходят для ответа на вопросы о различиях в некоторых статистических показателях (например, средних значениях, дисперсиях и т. д.), эти процедуры не подходят для решения вопросов, связанных с эквивалентностью. Затем, основная цель этой статьи состоит в том, чтобы сравнить наши недавно разработанные тесты эквивалентности групповых дисперсий с рекомендуемыми в настоящее время тестами однородности дисперсии в условиях данных, обычных для образовательных и психологических исследований. Обзор традиционных тестов однородности дисперсии, а также проверки эквивалентности намечается перед разработкой новых процедур проверки эквивалентности для обнаружения однородност

КОНТАКТЫ Роберт А. Крибби cribbie@yorku.ca Программа количественных методов, факультет психологии, Йоркский университет, Торонто, ON M3J 1P3, Канада.

Зачем проверять эквивалентность дисперсий?

Одной из наиболее распространенных причин, по которой исследователи хотят проверить эквивалентность групповых дисперсий, является необходимость обосновать использование тестов, предполагающих однородность дисперсии в их первичном анализе. В этом случае исследователь хотел бы найти, что дисперсии равны между группами. Важно отметить, что это нет необходимости использовать предварительный тест гомоскедастичности дисперсии, чтобы оправдать использование гетероскедастических процедур (например, гетероскедастический дисперсионный анализ Уэлча вместо традиционного дисперсионного анализа) потому что эти тесты, как правило, эффективны независимо от того, равны или неравны дисперсии между группы. Многие исследователи предлагали полностью отказаться от неробастных параметрических процедур. отдают предпочтение надежным процедурам, не требующим предположения об однородности дисперсий (например, Уилкокс, Чарлин и Томпсон, 1986; Циммерман, 2004). Однако исследователи в области педагогики и бихевиоризма по-прежнему широко используют традиционные параметрические процедуры и нуждаются в проверке допущений. связанных с этими испытаниями.

Более интересная причина для оценки эквивалентности дисперсий состоит в том, что основной исследовательский вопрос касается того, одинакова ли дисперсия зависимой переменной в нескольких группы. Как отмечает Парра-Фрутос (2009) , исследователи все больше интересуются свойствами их данных помимо центральной тенденции, такой как дисперсия или изменчивость. Например, все более распространенными становятся исследовательские вопросы, касающиеся «однородности» или «сходства» групп, включая вопросы о сопоставимости дисперсии оценок между группами. Брык и Рауденбуш (1988) утверждают, что наличие неоднородности дисперсии между группами может иметь важные последствия для выводы исследования. В частности, наличие неоднородности дисперсий в экспериментальном исследование указывает на наличие взаимодействия между характеристиками человека и лечебной группой членство. Другими словами, неоднородность дисперсий может указывать на то, что люди различаются по своим ответ на лечение (при условии, что в группе лечения был фиксированный эффект). Это может быть важным соображением для исследователей, и валидные тесты для оценки неоднородности или однородности дисперсий (в зависимости от ожиданий исследователя) важно оценивать в Экспериментальная дизайн. Действительно, в более сложных процедурах моделирования сравнение изменчивости, связанной с конкретным эффектом (например, изменчивость вокруг точки пересечения или наклона кривой скрытого роста модель) между разными группами является общей целью исследования (например, нет различий между группы по изменчивости вокруг склона).

Учитывая эти две причины для проверки однородности дисперсии, действительный тест для оценки эквивалентности дисперсии весьма актуален для вопросов, которые задают исследователи в области образования (и исследователи в смежные дисциплины) интересны и необходимы, если исследователь хочет обосновать использование традиционной тест средней разницы. Однако, как мы утверждаем в этой статье, имеющиеся в настоящее время процедуры неправильно оценивают равенство дисперсий, поэтому необходимо разработать и оценить новые процедуры.

Традиционные подходы к тестированию однородности дисперсии

Чтобы оценить однородность дисперсии, Левен (1960) предложил преобразовать баллы выборки в абсолютные отклонения выборочных баллов от выборочного среднего с zij D j Xij ¡ Mj j , где Xij оценка i-го индивидуума в j-й группе, а Mj — среднее значение j-й группы, а затем использование традиционного F-критерия ANOVA для zij для оценки равенства дисперсий по группам. Нулевая гипотеза для процедуры Левена состоит в том, что дисперсии совокупности всех J групп равны, H0 : s2 1 Д с2 2 Д … Д с2 Дж .
Альтернативная гипотеза утверждает, что по крайней мере одна групповая дисперсия не равна по крайней мере одной другой.

С тех пор как тест Левена был опубликован, было предложено множество модификаций. исходная версия демонстрирует некоторые нежелательные статистические свойства, такие как низкая мощность по сравнению с другие тесты (особенно при неравных размерах выборок) и неустойчивость к ненормально распределенным Xij. Предыдущие исследования моделирования (например, Conover, Johnson, & Johnson, 1981; Keselman, Games, & Clinch, 1979 год; Лим и Лох, 1996; Nordstokke & Zumbo, 2010) дали ряд рекомендаций. относительно оптимального теста однородности дисперсии, который также устойчив к ненормальности. Например, Коновер и др. (1981) предполагают, что исходный критерий Левена с использованием медианы является одним из наиболее эффективных статистических показателей в широком диапазоне аналитических условий. Лим и Лох (1996) также рекомендуют

Levene тестирует с использованием медианы, но предполагает, что версия с начальной загрузкой улучшает производительность этой статистики. Нордстокке и Зумбо (2010) рекомендовали ранговый тест Левена как наиболее надежную статистику теста для многих условий данных, а ранговые тесты Левена также были рекомендованы в исследовании Con over как обладающие некоторыми желательными свойствами при определенных условиях. Кесельман и др. (1979) сообщают, что ни один из тестов не может быть рекомендован единообразно, поскольку эффективность многих статистических показателей дисперсии-однородности зависит от условий анализа. Однако они предположили, что исходный Levene, использующий медиану, или Levene, использующий медиану с поправкой Welch, может быть л В более позднем исследовании Кесельман, Уилкокс, Альгина, Отман и Фрадетт (2008) рассмотрели стратегии с усеченными средними и предположили, что первоначальный Levene с усеченными средними или Levene с использованием усеченных средних с поправкой по Уэлчу показал лучшие результаты в оцениваемых условиях. (только на основе частоты ошибок типа I). Эти авторы также предполагают, вопреки исследованию Лима и Ло, что бутстреп не был необходим, потому что удовлетворительная частота ошибок типа I может быть получена без бутстрэпа. Несмотря на почти 50 лет исследований, похоже, не существует общего консенсуса в отношении единого статистического теста для оценки однородности дисперсий, который одинаково хорошо работал бы в сценариях с общими данными. Однако тест Левена, основанный на медиане, часто рекомендуется, поскольку он хорошо работает в широком диапазоне условий.

Традиционные процедуры однородности дисперсии, оцениваемые в текущем исследовании В текущем

исследовании оценивались четыре традиционных теста на основе различий на однородность дисперсий, каждый из которых описан ниже.

Оригинальный тест Левена (1960) на однородность дисперсий («Lev_mean»)

Хотя тест Levene (1960) не рекомендовался в литературе (например, Conover et al., 1981; Lim & Loh, 1996), он все еще регулярно используется в популярных статистических программах, поэтому он был включен в данное исследование.

Тест Левена с использованием медианы («Lev_mdn»)

Эта модификация теста Левена, первоначально предложенная Брауном и Форсайтом (1974), считалась наилучшей процедурой в имитационном исследовании Коновера и др. (1981) с точки зрения наиболее точной частоты ошибок типа I. Вместо использования среднего значения j-й выборки в преобразовании оценки выборки эта модификация использует преобразование ~zij D j Xij ¡ MDNj j , где MDNj — медиана j-й выборки. Преобразованные баллы анализируют с использованием F-теста ANOVA.

Оригинальный тест Левена с поправкой Уэлча («LevWelch_mean»)

Процедура Welch (1951) с поправкой на гетероскедастические степени свободы была предложена как решение проблемы неравной дисперсии в процедурах проектирования независимых групп, таких как t-критерий Стьюдента и F-критерий ANOVA. Однако поправка Уэлча к F-критерию дисперсионного анализа также имеет отношение к критерию Левена на однородность дисперсий (и его модификациям), учитывая, что критерий Левена использует критерий дисперсионного анализа F и, таким образом, также предполагает однородность дисперсий (точнее, однородность дисперсий абсолютные значения баллов отклонений, zij). Кажется нелогичным иметь критерий однородности дисперсий, который сам по себе предполагает однородность дисперсий. Таким образом, исследователи предложили использовать статистику, скорректированную по Уэлчу, для проверки однородности дисперсий (например, Keselman et al., 1979; Parra-Frutos, 2009; Wilcox et al., 1986).

Как и в исходном тесте Левена, преобразованные значения zij D j Xij ¡ Mj j просто подставляются в уравнение F' для оценки однородности дисперсий (не требуя однородности дисперсий).

предположение дисперсии), так что статистика теста становится:

Ф° Д Xwzj .Zj ¡ Z ::/ 2 6 Дж ¡ 1 / 1 С 2.J ¡ 2/ J2 ¡ 1 x 1 / нджj ¡ 1 1 ¡ wzj / Иксwzj !2 ;

где wzj D c2 zj, nj – размер j-й группы, с 2 zj — дисперсия преобразованных оценок для j-го группа, З Д Пвж Zj / п wzj, Zj — среднее значение zij для j-й группы. Ф ° статистика примерно распределяется как F со степенями свободы в числителе J-1 и степенями свободы в знаменателе:

дф° Д J2 ¡ 1 / 3 x 1 / нджj ¡ 1 1 ¡Xwzj wzj 2 :

Медианный критерий Левена с поправкой Уэлча («LevWelch_mdn»)

В этой процедуре используются абсолютные отклонения от медианы, ~zij D j Xij ¡ MDNj j , провести дисперсионный анализ Уэлча F0 для оценки однородности дисперсий (описанных ранее). В этом случае Zj является средним zij для j-й группы . Учитывая, что версия процедуры Брауна-Форсайта наиболее широко рекомендуется в литературе, в данное исследование была включена версия этого теста Уэлча.

Задачи с традиционными тестами эквивалентности дисперсий

Несмотря на то, что результаты предыдущих исследований моделирования показали, что ряд однородных дисперсионных тестов адекватно работает при различных условиях данных, все они в корне неверны. для проблемы определения равенства дисперсий совокупности, поскольку эти основанные на различиях процедуры направлены на то, чтобы не отвергнуть нулевую гипотезу относительно точного равенства дисперсий групп. В частности, если используется традиционный тест на однородность дисперсий, цель состоит в том, чтобы не отклонить нулевое значение. гипотеза для этих тестов, H0: s2 1 Д c2 2 Д ... Д c2 Дж . Другими словами, исследовательская гипотеза о равенстве дисперсий согласуется с нулевой гипотезой, а не с альтернативной гипотезой. Вероятность ошибки I рода при проверке нулевой гипотезы H0: s2 1 Д c2 2 Д ... Д c2 Дж это шанс неправильно сделать вывод, что существует разница между дисперсиями, когда на самом деле нет различий в отклонения. Контроль частоты ошибок типа I — это защита от неправильного определения различий между двумя или более отклонений, когда они одинаковы. Однако, если не удается отвергнуть истинную нулевую гипотезу, не может заключить, что дисперсии эквивалентны; неспособность отвергнуть нулевую гипотезу H0: s21 Д c2 2 Д ... Д c2 Дж означает только то, что нет достаточных доказательств, чтобы сделать вывод о наличии различий между отклонения.

Другая проблема с традиционными тестами заключается в том, что отклонение или неотвержение нулевой гипотезы об однородности дисперсии очень мало говорит о потенциальном сходстве рассматриваемых групповых дисперсий. В частности, нулевая гипотеза, оцениваемая с помощью критериев однородности дисперсии на основе различий, слишком специфичны и непрактичны для оценки эквивалентности групповых дисперсий. Например, если есть большой размер выборки и очень незначительная разница между групповыми дисперсиями, вполне вероятно, что тест на однородность дисперсии, основанный на различиях, отвергнет нулевую гипотезу и объявит дисперсии генеральной совокупности неравный. Однако обычно ожидаются небольшие различия в дисперсиях, и, таким образом, результаты традиционная однородность теста дисперсии и последующие выводы о сходстве дисперсии населения в этом случае могут быть нецелесообразными. И наоборот, меньшие размеры выборки могут привести к очень мало возможностей для обнаружения важных различий в дисперсиях, что приводит к неточным выводам о различиях населения. В целом, традиционные процедуры, основанные на различиях, менее вероятны. для обнаружения равенства дисперсий по мере увеличения размера выборки, что несовместимо с типичными ожиданиями проверки нулевой гипотезы.

Проверка эквивалентности

Эквивалентные тесты подходят для исследовательского вопроса, связанного с отсутствием ассоциации. За Например, исследователя может заинтересовать демонстрация того, что средства групп эквивалентны, что между двумя переменными не существует связи, эти переменные не взаимодействуют (например, Крибби, Груман и Арпин-Крибби, 2004 г.; Крибби, Рагунанан и Коунселл, 2016 г .; Герцен и Крибби, 2010 г.; Робинсон, Дуурсма и Маршалл, 2005 г .; Роджерс, Ховард и Весси, 1993 г., Шуирманн, 1987 г .; Веллек, 2010), или что дисперсии двух или более популяций равны (как предложено в текущем исследовании). При проверке эквивалентности отсутствие связи подразумевает, что разница между двумя статистическими данными настолько мала, что может считаться бессмысленным или бессмысленным. Эта разность определяется априори как эквивалентность интервал (например, ¡d, d; более подробно обсуждается позже). Другими словами, тест эквивалентности оценивает, является ли связь между двумя или более объектами (например, разница между дисперсиями генеральной совокупности) падает в пределах заданного интервала, определяющего несущественное различие (например, ¡ d $s2_1$ ¡ $c2_2$ г).

Новые тесты на основе эквивалентности и однородности дисперсии, оцененные в текущем исследовании

Учитывая фундаментальные проблемы с традиционными тестами на однородность дисперсий, мы разработали основанный на эквивалентности тест на однородность дисперсий вместе с несколькими модификациями. Ранее, Wellek (2010) разработал подход к оценке эквивалентности дисперсий для двух групп, которые использует отношение наибольшей дисперсии к наименьшей, что, как предложил анонимный рецензент, по своему характеру аналогичен тесту Fmax , разработанному Хартли (1950). В текущем исследовании изучается альтернативный подход, использующий необработанную разницу, а не отношение, и подходящий для двух или более независимых групп. Нулевая гипотеза для одностороннего теста эквивалентности на однородность дисперсий: что разница между дисперсиями групп равна или больше априорно определенной отсечка (e^2):

$$H0 : C^2 \quad e2$$
$$H1 : C^2 < e2$$

где C^2 количественно определяет разницу между дисперсиями групп и e^2 представляет собой наименьший разница в дисперсиях, которая считается значимой. Обратите внимание, что интервал эквивалентности включает любое значение меньше e^2 и содержит нижнюю границу нуля, так как мы работаем в квадратах единиц (которые отличается от интервала эквивалентности многих других тестов эквивалентности, симметричных относительно нуля). Больше обсуждения относительно e^2 приведен ниже.

Критерий Левена-Веллека на эквивалентность дисперсий («LW_mean»)

Эта процедура основана на исходной статистике однофакторного теста эквивалентности Wellek (2010) (которая одновременно оценивает эквивалентность всех средних значений J-популяции), заменяя необработанные оценки исходным преобразованием Левена. Эта новая гибридная тестовая статистика может быть представлена как:

$$ДВ\ Д \frac{\sum_{Дж\ Д\ 1}^{Дж} nj\ \overline{Zj}\ ¡\ \overline{Z}^{2} \quad 6\ Дж\ ¡\ 1}{\sum_{я\ Д\ 1}^{Xn} \sum_{Дж\ Д\ 1}^{Дж} .zij\ ¡\ \overline{Zj}\ /^{2} \quad 6\ .Н\ ¡\ Дж/}$$

с исходным преобразованием Левена, zij D j Xij ¡ Mj j , так что Zj D и Z — Pzij — Пиджей Дж Д 1 пНЖя Д 1зий / PJ j D

ne2 , H0: С где e2 вычисляется, если LW умножен на Дж ¡ 1; Н Дж — среднего размера ^{2} 1nj представляет собой параметр нецентральности, группы на квадрат верхней границы интервала эквивалентности. Важно отметить, что для простоты мы представили LW как F-статистику, в отличие от традиционной формулировки в метрике c2 (Wellek, 2010).

Как упоминалось ранее, как исходный тест Левена, так и односторонний тест Веллека предполагают однородность дисперсий, что является необоснованным предположением, когда эти тесты используются для оценки однородности дисперсий. Кроме того, предыдущие исследования традиционных тестов на однородность дисперсии на основе различий показали, что некоторые модификации исходного теста Левена работают лучше. Таким образом, это исследование включало три дополнительные процедуры, основанные на модификациях этого недавно разработанного теста Левена-Веллека, как описано ниже.

Левене-Веллек с использованием медианы («LW_median»)

Эта процедура является адаптацией критерия Левена-Веллека (определенного выше) с использованием абсолютных отклонений от медианы выборки вместо абсолютных отклонений от среднего значения выборки (т. е. преобразование Брауна-Форсайта выборочных оценок).

Левен-Веллек-Уэлч («LWW_mean»)

Эта версия процедуры основана на критерии Левена-Веллека для среднего, но включает поправку Уэлча для проверки эквивалентности групповых дисперсий без предположения об однородности (преобразованной оценки) дисперсий. Новая статистика надежного теста, основанная на эквивалентности, может быть представлена как:

$$LWW\ Д \frac{Xwzj\ .Zj\ ¡\ Z\ 1\ ¡Xwzj\ ::/^{2}}{1\ С\ 2.J\ ¡\ 2/ \ J2\ ¡\ 1\ x \ \frac{1}{ндж\ ¡\ 1}\ wzj\ ^{2}} \ \frac{Дж\ ¡\ 1}{Н}$$

Как и в случае с LW-критерием, тестовая статистика примерно распределена как нецентральная F с нонценом. параметр верности $\bar{ne}^{2}$, J - 1 степень свободы в числителе и степени свободы в знаменателе df ^{0}.

Левене-Веллек-Уэлч с использованием медианы («LWW_median»)

Последняя новая процедура, разработанная для этого исследования, использует ранее определенный критерий Левена-Веллека-Уэлча, но вместо первоначального преобразования Левена в этой процедуре используется преобразование Брауна-Форсайта абсолютных отклонений выборочных оценок от медианы.

Интервал эквивалентности Wellek

(2010) дает несколько общих рекомендаций по выбору интервалов эквивалентности.
Однако определяющим фактором при выборе соответствующего интервала эквивалентности должен быть характер исследования. В самом деле, Веллек и другие исследователи, занимающиеся тестированием эквивалентности, предупредили, что общие рекомендации или фиксированные общие правила относительно выбора интервала эквивалентности не рекомендуются, но должны быть предметом тщательного рассмотрения, относящегося к конкретному исследова
Эпсилон (e) можно описать как максимальную разницу в дисперсиях, которую можно было бы считать несущественной.
В целом Веллек предполагает, что объекты, отличающиеся не более чем на 10%, очень похожи, а различия более чем на 20% практически значимы. Таким образом, разница в 10 % будет строгим критерием эквивалентности (e D .25), а 20 % будет более либеральным критерием эквивалентности (e D .50; подробности см. Wellek, 2010, стр. 16, 17, 22). .

Вопросы, которые следует учитывать при сравнении тестов эквивалентности и тестов различий

Важно обсудить некоторые трудности, связанные со сравнением результатов тестов, основанных на различиях, с результатами тестов, основанных на эквивалентности. Основная проблема заключается в том, что эти два типа тестов оценивают разные гипотезы. Тесты на основе различий оценивают точечно-нулевую гипотезу, которая очень специфична, а в случае равенства дисперсий совершенно непрактична. Например, абсолютно невозможно установить, что дисперсии генеральной совокупности в точности равны, если использовать достаточно десятичных разрядов. Кроме того, исследовательская гипотеза о равенстве дисперсий согласуется с нулевой гипотезой, а не с альтернативной гипотезой, поэтому цель исследователя состоит в том, чтобы «принять» нулевую гипотезу.

Тесты на основе эквивалентности оценивают нулевую гипотезу о том, что разница между дисперсиями выходит за пределы заранее заданного интервала эквивалентности. Таким образом, чтобы определить, что дисперсии почти эквивалентны, нужно отвергнуть эту нулевую гипотезу и вместо этого найти, что разница между дисперсиями попадает в интервал эквивалентности. В этом случае исследовательская гипотеза является альтернативной гипотезой, которая соответствует обычным процедурам проверки нулевой гипотезы. Тем не менее, можно провести сравнение общей картины результатов для выявления однородности различий между этими двумя методами тестирования. Результатом этого исследования стала доля заявлений об эквивалентности. Другими словами, какова была вероятность обнаружения эквивалентности («мощность» обнаружения эквивалентности)? Этот результат определялся долей неотвержений нулевой гипотезы в тестах, основанных на различиях, и долей отклонений нулевой гипотезы в тестах, основанных на эквивалентности.

Метод

Моделирование Монте-Карло использовалось для сравнения вероятности объявления эквивалентности для четырех основанных на различиях тестов на однородность дисперсий с вероятностью для четырех новых основанных на эквивалентности тестов на равенство дисперсий. Кроме того, оценивались и сравнивались частота ошибок типа I и мощность процедур эквивалентности. Эффективность восьми тестов на однородность дисперсии оценивалась с использованием нормального распределения населения и распределения населения с положительной асимметрией (x2 с тремя степенями свободы). Чтобы оценить частоту ошибок типа I процедур, основанных на эквивалентности, были использованы либеральные оценки § 0.5a (Bradley, 1978) . Таким образом, при уровне 0,05 считалось, что процедура имеет точную эмпирическую частоту ошибок типа I в конкретных условиях, если частота находилась в диапазоне от 0,025 до 0,075. Важно отметить, что если для конкретного условия получена неточная частота ошибок типа I, соответствующий коэффициент мощности для этого условия также следует интерпретировать с осторожностью, поскольку он может быть искусственно завышен или занижен в результате неточной ошибки типа I. контроль ошибок. Моделирование проводилось с помощью статистического программного обеспечения R с открытым исходным кодом (R Core Team, 2016).

Определение мощности для тестов, основанных на эквивалентности, отличается от тестов, основанных на различиях, потому что, как обсуждалось ранее, эти два типа тестов имеют разные нулевые гипотезы. Поэтому вместо определения вероятности отклонения ложной нулевой гипотезы, то есть «мощности» для любого конкретного теста, в данном исследовании определялась «вероятность нахождения эквивалентности» как для процедур, основанных на эквивалентности, так и для процедур, основанных на различиях. Другими словами, это исследование было сосредоточено на вероятности того, что конкретный тест объявляет дисперсии эквивалентными, когда они фактически эквивалентны (где эквивалентность определяется нулевой гипотезой для тестов, основанных на различиях, и альтернативной гипотезой для тестов, основанных на эквивалентности).). Эмпирические коэффициенты ошибок типа I для тестов, основанных на эквивалентности, были получены путем определения различий в дисперсиях, которые соответствовали границам интервала эквивалентности (т . е. C2 D e2) в условиях, когда дисперсии генеральной совокупности различались между группами. См. Рисунок 1 для меха .
там уточнение.

Мы рассмотрели группы JD 4 для всех условий. В этом исследовании манипулировали несколькими переменными, включая размер выборки (n = 10, 25, 50, 100), сбалансированный и несбалансированный планы (т. е. равные или неравные размеры групп), равенство/неравенство дисперсий генеральной совокупности и сочетание неравных размеров выборок с неравные дисперсии населения. Сводку условий, проверенных в этом исследовании, можно найти в таблице 1. Для тестов на основе эквивалентности использовались консервативный и либеральный пределы эквивалентности e D .25 и e D .50 соответственно (Wellek, 2010). Тем не менее, картина результатов для эквивалентности

Machine Translated by Google

		НАСЕЛЕНИЕ	
		Эквивалент	Не эквивалентно
ОБРАЗЕЦ	Эквивалент	Дифф - правильное решение	Дифференциал — ошибка типа II
		Эквивалент - Мощность	Эквивалент — ошибка типа I
	Нет Эквивалент	Дифференциал — ошибка первого рода	Дифференциал - Мощность
		Эквивалент – ошибка второго рода	Эквивалент – правильное решение

Рис. 1. Описание различий в определении мощности, ошибок типа I и ошибок типа II между тестами на различия и тестами на эквивалентность. Для тестов, основанных на различиях, Эквивалент подразумевает, что все отклонения идентичны; тогда как для основанного на эквивалентности тестов, Эквивалент подразумевает, что разница в дисперсиях меньше минимальной значимой разницы. Кроме того, для тестов, основанных на различиях, Not Equivalent подразумевает любую разницу в дисперсиях; тогда как для тестов на основе эквивалентности Not Equivalent подразумевает разницу, большую или равную минимальной значимой разнице. Diff D тест на основе различий. Тест эквивалентности Equiv D.

пределы были одинаковыми, поэтому представлены только результаты для е D .50. Как и ожидалось, тарифы на электроэнергию для е D .25 были ниже во всех условиях.

Для нормально распределенных условий было сгенерировано nj стандартных нормальных наблюдений для j-го группу, где j D 1, ..., J, а полученные значения умножались на sj так, чтобы наблюдения есть отклонения, s2 дж, как указано в Таблице 1. Чтобы изучить влияние положительно асимметричного распределения На выполнение тестовой статистики было сгенерировано nj наблюдений для каждой из J групп. из распределения x2 с тремя степенями свободы. Для того, чтобы убедиться, что наблюдения от Распределение x2 имело дисперсию, указанную в таблице 1, сначала среднее значение и дисперсию распределения. должны быть установлены на 0 и 1 соответственно. Это было достигнуто путем вычитания среднего (среднее D df D 3) и деление на стандартное отклонение (результирующие 2df p D ð Þ2 ð Þ3 p 2.45) распределения x2 . значения SD D затем умножались на sj для получения распределения наблюдений с дисперсиями изложены в таблице 1.

Несбалансированные планы (т. е. неравные размеры выборки в группах), которые сочетаются с неравными дисперсиями, могут серьезно повлиять на контроль ошибок типа I и типа II в процедурах типа ANOVA (Кесельман). и др., 1998; Отман и др., 2004). Таким образом, в настоящем исследовании изучались как положительные (прямые), так и отрицательные (обратные) пары дисперсий и размеров выборки. Положительное спаривание происходит, когда самая большая группа размер сочетается с наибольшей дисперсией, а наименьший размер группы сочетается с наименьшей дисперсией. Отрицательное спаривание происходит, когда наибольший размер группы сочетается с наименьшей дисперсией, а наименьший размер группы сочетается с наибольшей дисперсией. Предыдущие исследования надежности дисперсионного анализа типа Процедуры (Othman et al., 2004; Yin & Othman, 2009) обнаружили, что положительные пары приводят к консервативной частоте ошибок типа I, а отрицательные пары приводят к либеральной частоте ошибок типа I. Пример пары размеров можно найти в таблице 1.

Как только наблюдения были созданы для каждой репликации, четыре процедуры, основанные на различиях, и четыре процедуры на основе эквивалентности были выполнены на данных каждой повторности. Чтобы определить вероятность объявления эквивалентности для тестов, основанных на различиях, было отмечено, когда нуль

Таблица 1. Интервалы эквивалентности (е), дисперсии генеральной совокупности (s2), формы распределения (λ) и размеры выборки (n) для имитационного исследования.

	s2 (λ D нормальный)		s2 (λ D x2 3 df)	
Тип условия	е Д .25	е D .50	е Д .25	е D .50
Ошибка типа I	1, 1,22, 1,45, 1,67 1, 1,	1, 1,64, 2,28, 2,92 1, 1,	1, 1,28, 1,56, 1,84 1, 1,	1, 1,85, 2,70, 3,55
Сила	1, 1 1, 3, 4, 6 1,	1, 1 1, 3, 4, 6 1,	1, 1 1, 3, 4, 6 1,	1, 1, 1, 1
Ошибка типа I	1,1, 1,2, 1,3	1,33, 1,66, 2	1,1, 1,2, 1,3	1, 3, 4, 6
Сила				1, 1,33, 1,66, 2
	Условия размера выборки			
	п Д 10	п Д 25	п Д 50	н Д 100
Равные размеры образцов	10, 10, 10, 10 5,	25, 25, 25, 25 18,	50, 50, 50, 50 25,	100, 100, 100, 100
Положительные пары	8, 12, 15 15, 12,	22, 28, 32 32, 28,	40, 60 75 75, 60,	50, 80, 120, 150
Отрицательные пары	8, 5	22, 18	40, 25	150, 120, 80, 50

гипотеза не была отвергнута. Для определения вероятности объявления эквивалентности проверки на основе эквивалентности (т. е. мощности) было отмечено, когда нулевая гипотеза была отвергнута. Этот процесс был повторен в 10 000 повторений для каждого условия, чтобы получить вероятность объявления эквивалентности для каждого условия.

Полученные результаты

Как отмечалось ранее, из-за сходства результатов представлены только результаты для е D 0,50.1

Процедуры на основе эквивалентности

Эмпирическая частота ошибок типа I

Сводные данные о частоте ошибок типа I можно найти в таблице 2; напомню, что это ошибка первого рода полученные при составлении сводки процедур [illegible] ниже.

Нормальные распределения. Частота ошибок типа I в условиях одинакового размера выборки поддерживалась близкой к номинальный уровень в диапазоне от 0,0381 до 0,0702. Для положительных условий спаривания процедуры Levene-Wellek Welch (LWW_mean, LWW_median) имели приемлемую частоту ошибок типа I для всех размеров выборки. Однако обе процедуры Левена-Веллека (т. е. LW_mean, LW_median) имели слишком либеральный тип. Частота ошибок I при максимальном размере выборки (0,084 и 0,0869). Для условий отрицательного спаривания LWW_median имел очень либеральную частоту ошибок типа I при наименьшем размере выборки (0,1014 при n D 10). Однако при больших размерах выборки в условиях отрицательного спаривания частота ошибок типа I была ниже. приемлемо для LWW_median. Остальные три процедуры эквивалентности сохранили ошибку первого рода. ставки в пределах от 0,025 до 0,075 во всех отрицательных условиях спаривания.

Распределения с положительной асимметрией ($x2$, 3 df). Все процедуры эквивалентности поддерживаются точными. Тип Частота ошибок I, когда дисперсии были отрицательно связаны с неравными размерами выборки. Однако, когда дисперсии положительно сочетались с наибольшим неравным размером выборки, n = 100, только LWW_медиана имела значение. точная частота ошибок типа I. Кроме того, LWW_mean имел частоту ошибок типа I, которая была слишком консервативной (0,0196), когда n = 10 и размеры выборки были положительно связаны с дисперсиями и типом I. частота ошибок, которая была слишком либеральной (0,0814) в условиях наибольшего равного размера выборки.

Сила

Сводку результатов по мощности для процедур эквивалентности можно найти в таблице 3. Когда дисперсии были точно равными, разница в дисперсиях (т. е. нулевая разница) попадала в пределы эквивалентности. интервал, так что это было условие мощности для процедур, основанных на эквивалентности. Кроме того, для 2:1

Таблица 2. Сводная информация о частоте ошибок типа I: минимальная и максимальная эмпирическая частота ошибок типа I и количество раз, превышающее частоту ошибок типа I. превысило пределы 0,025–0,075 для процедур эквивалентности по 24 нулевым условиям (C2 D e^2).

Тест	Минимальный эмпирический Частота ошибок типа I	Максимальная эмпирическая Частота ошибок типа I	Количество ошибок типа I Частота ошибок Превышены границы 0,025–0,075
Левене-Веллек иметь в виду	0,0228	.1113	4
Левене-Веллек медиана	0,0223	0,0869	3
Левене-Веллек валлийское среднее	0,0196	0,0850	3
Левене-Веллек валлийская медиана	0,0356	.1014	1

В эту таблицу не включены условия, при которых C2 > e^2.

Таблица 3. Сводка мощности: количество условий (из 48 условий), в которых конкретная процедура эквивалентности имела наивысшее значение. мощность (т. е. условия, при которых нулевая гипотеза была ложной).

Тест	Тест имел наивысшую мощность в Equal Условия размера выборки (из 16)	Тест имел наибольшую мощность в положительном результате Условия сопряжения (из 16)	Тест имел наивысшую мощность в отрицательном результате Условия сопряжения (из 16)
Левене-Веллек иметь в виду	0 00		
Левене-Веллек медиана	15	7	6
Левене-Веллек валлийское среднее	0 00		
Левене-Веллек валлийская медиана	1	6	10

Исключая условия, при которых была ничья для наилучшей процедуры.

отношение дисперсии эта разница в дисперсиях находилась в пределах интервала эквивалентности для эквивалентности процедуры, такие что $c^2 < e^2$; поэтому это условие было еще одним испытанием силы этих процедуры.

Нормальные распределения. Мощность более 90% для обнаружения эквивалентности была достигнута, когда $\bar{n}$ было равно 50. и достиг почти 100% в условиях наибольшего размера выборки. Этот результат получен для равной выборки размеры и условия положительного и отрицательного спаривания.

Для коэффициента дисперсии 2:1 в условиях наибольшего размера выборки мощность составляла примерно от 41% до 61%. Все четыре процедуры эквивалентности имели сопоставимые мощности для всех комбинаций размера выборки и дисперсии в условиях нормального распределения.

Распределения с положительной асимметрией (χ^2, 3 df). Мощность приблизилась к 99% для процедур, основанных на медиане. когда дисперсии были точно равны. Однако для процедур, основанных на среднем значении, мощность была немного ниже, примерно на 95%.

Для коэффициента дисперсии 2:1, когда размеры выборки были одинаковыми, процедуры, основанные на медиане, имели самые высокие значения. мощность при всех размерах выборки. Это преимущество в мощности для тестов, основанных на медиане, также наблюдалось, когда неравные размеры выборки были положительно связаны с дисперсией, а когда неравные размеры выборки были отрицательно связаны с дисперсией. См. рис . 2.

Ложные декларации эквивалентности

Для коэффициента дисперсии 6:1 C^2 было больше, чем e^2; Таким образом, различия в дисперсиях превысили интервал эквивалентности и процедуры эквивалентности не должны отвергать нулевую гипотезу дисперсии неоднородность. Это также было еще одной оценкой частоты ошибок типа I процедур эквивалентности. при условии, что нулевая гипотеза о неоднородности дисперсии была верна в этом состоянии. В частности, разница между групповыми дисперсиями превышала интервал эквивалентности. Обратите внимание, однако, что ошибка частоты в этом условии отношения дисперсии должны быть меньше, чем частоты ошибок типа I, полученные, когда различия между дисперсиями совпадают с границами интервала эквивалентности.

Нормальные распределения. Как и ожидалось, вероятность объявления эквивалентности была низкой при небольшой выборке. размеров и был равен нулю в условиях большего размера выборки.

Распределения с положительной асимметрией (χ^2, 3 df). Частота ошибок была почти нулевой при размерах выборки. равные или когда неравные размеры выборки были положительно связаны с дисперсиями. Частота ошибок была несколько выше для условий отрицательного спаривания, хотя они оставались близкими к эмпирическому Типу Я частота ошибок. В условиях самых больших размеров выборки частота ошибок во всех условиях была равна нулю или почти нулевой (см. рис . 3).

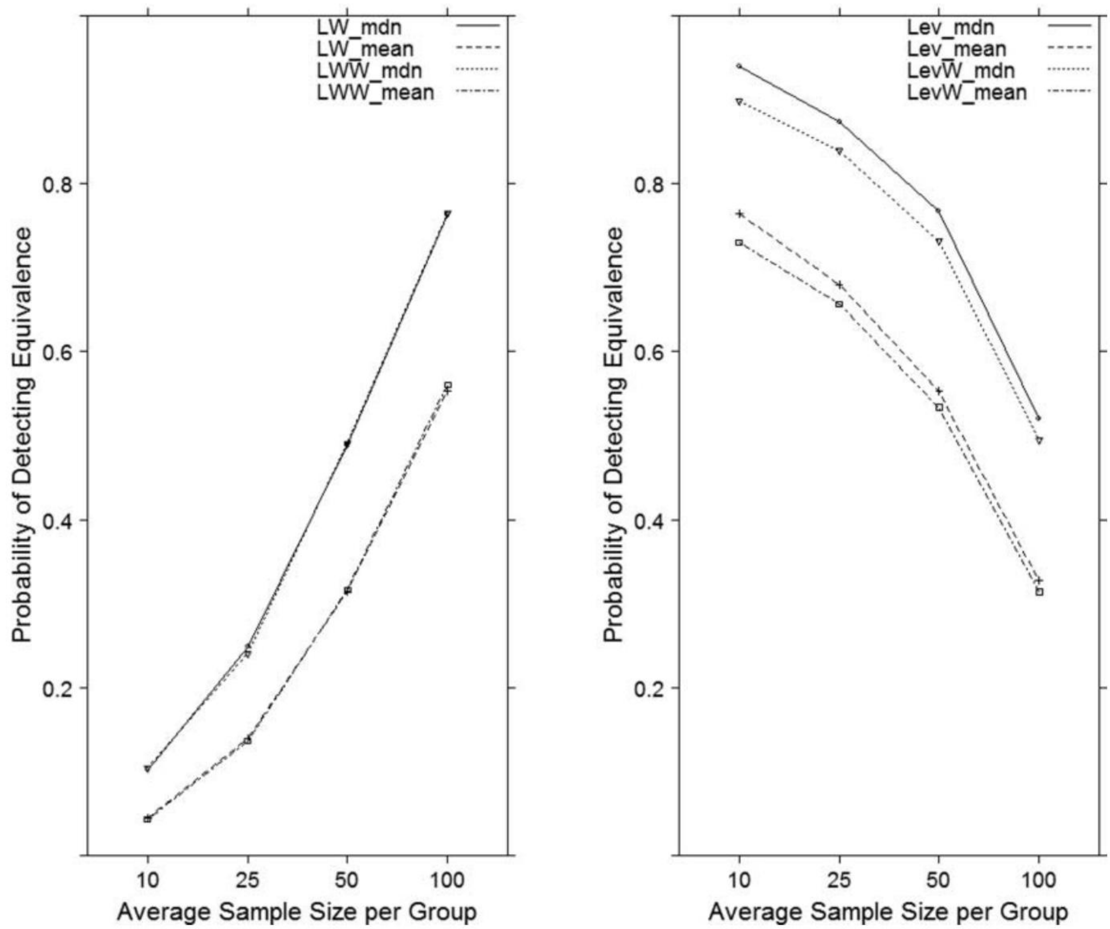

Рис. 2. Средняя вероятность объявления эквивалентности; e D 0,50, s2 D 1, 1,33, 1,66, 2 (C2 ^{2}); x2 раздачи; Левая панель D эквивалент <процедуры e lence; правая панель D, основанные на различиях процедур.

Процедуры на основе различий

Нормальные распределения. Когда дисперсии генеральной совокупности групп были точно равными, это было условием ошибки типа I для процедур, основанных на различиях. Следовательно, вероятность объявления эквивалентности (т. е. невозможности отвергнуть нулевую гипотезу) в этом условии должна была быть приблизительно 1 — а (в данном случае 0,95) — независимо от размера выборки. Хотя в большинстве случаев показатели были близки к 0,95, с положительными и отрицательными сочетаниями неравных размеров выборки и дисперсии и малых размеров выборки, показатели, как и ожидалось, были иногда слишком консервативными или слишком либеральными.

Для условия отношения дисперсии 2:1 тесты на основе различий имели очень высокую вероятность объявления эквивалентности при n D 10 (отметим, что это неверное решение, т. е. ошибка второго рода). В условиях наибольшего размера выборки (n = 100) вероятность объявления эквивалентности была намного ниже в условиях равного размера выборки. Важно отметить, что отношение дисперсии 2:1 в этом условии означало, что нулевая гипотеза процедур, основанных на различиях, была ложной, и, таким образом, эти результаты не были неожиданными. Тем не менее, обратная природа использования тестов, основанных на различиях, для решения вопросов эквивалентности была очевидна, поскольку эквивалентность обнаруживается до 97% времени при небольших размерах выборки, но та же самая разница в дисперсиях была статистически значимой большую часть времени в исследованиях. Условия наибольшего размера выборки.

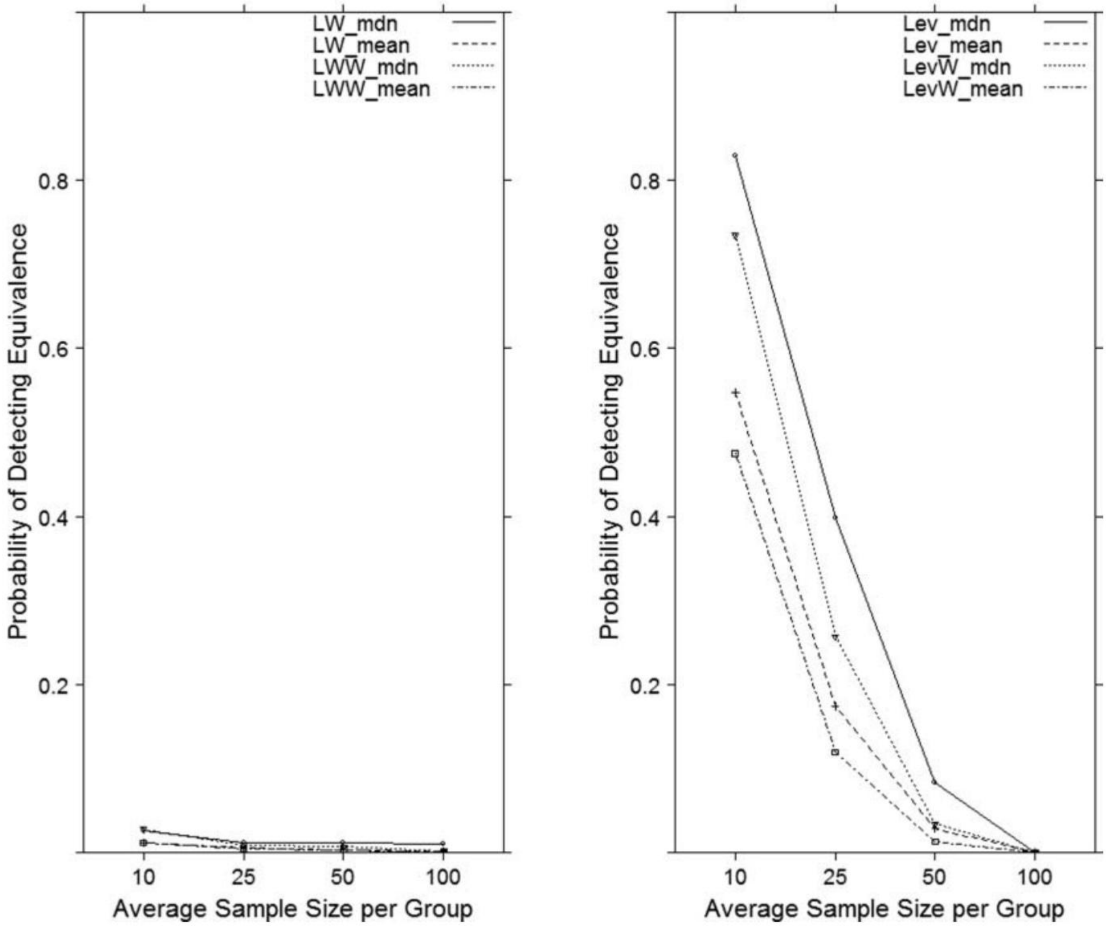

Рис. 3. Средняя вероятность объявления эквивалентности, e D .50, s2 D 1, 3, 4, 6 (C2 > e 2) и x2 распределения. Левая панель D эквивалент линзовые процедуры; правая панель D процедуры, основанные на различиях.

При соотношении дисперсии 6:1 в условиях наименьшего размера выборки вероятность объявления эквивалентности достигала 85% в условиях отрицательного спаривания и достигала 72% в условиях равного размера выборки.

Распределения с положительной асимметрией (x2 , 3 df). Как обсуждалось ранее, когда дисперсии групп были точно равными, это условие оценивало частоту ошибок типа I для процедур, основанных на различиях. Поэтому вероятность объявления эквивалентности в этом условии должна была быть примерно 1 — a (.95) для разностных процедур. Этот результат был получен для большинства повторений с тесты, основанные на медиане, но процедуры, основанные на среднем, продемонстрировали частоты, которые часто были очень консервативными. Показатели по процедурам варьировались примерно от 95% при тесте Левена с использованием в среднем, но были ниже 80% для других процедур. Таким образом, вероятность объявления эквивалентности меньше, чем в условиях нормального распределения. Обратите внимание, что, как и прежде, размер выборки не повлияло на вероятность объявления эквивалентности в этом условии для тестов, основанных на различиях.

Когда имелось отношение дисперсии 2:1, опять же, гипотеза нулевой точки для процедур, основанных на различиях, была ложной. Следовательно, вероятность объявления эквивалентности (т. е. не отбрасывания нулевого гипотеза) уменьшалась по мере увеличения размера выборки. См. рис . 2.

Для коэффициента дисперсии 6:1 наблюдалось много ложных заявлений об эквивалентности для процедур, основанных на различиях, в условиях меньшего размера выборки (см. рис . 3). Однако в условиях наибольшего размера выборки скорость была приблизительно равна нулю.

Обсуждение

Результаты имитационного исследования продемонстрировали отсталость традиционного метода, основанного на различиях. процедуры оценки равенства дисперсий генеральной совокупности. В частности, мощность для обнаружения эквивалентности был в неправильном направлении, так что увеличение размера выборки привело к снижению мощности для обнаружения эквивалентность дисперсий. Кроме того, результаты моделирования помогли продемонстрировать, что традиционная нулевая гипотеза непрактична, что важно, поскольку небольшие различия в дисперсиях часто несущественные и ожидаемые. Несмотря на то, что тесты, основанные на различиях, часто не нулевая гипотеза, когда были небольшие различия в дисперсиях, потому что они не были правильно выполнять. По мере увеличения размеров выборки увеличивались шансы объявить небольшие различия в дисперсиях важными различиями. И наоборот, большие и, возможно, важные различия в групповые дисперсии часто объявлялись незначимыми в тестах, основанных на различиях, когда размеры выборки были маленькими.

Учитывая эти проблемы с традиционными процедурами, основанными на различиях, процедуры, основанные на эквивалентности, более подходят, если целью исследования является оценка равенства дисперсий. Эквивалентные тесты согласовать исследовательскую гипотезу о равенстве дисперсий с альтернативной гипотезой, чтобы обнаружить эквивалентность и отклонить нулевую гипотезу увеличивается с размером выборки, как и ожидалось при использовании процедуры проверки нулевой гипотезы. Кроме того, использование интервальной гипотезы, а не точечная нулевая гипотеза позволяет исследователям диктовать, насколько или насколько мало перекрываются дисперсии может быть важным. В целом, небольшие различия в отклонениях ожидаются и обычно не имеет значения, поэтому тест, предназначенный для оценки примерного равенства, гораздо более практичен, чем тесты которые оценивают точную эквивалентность (т. е. нулевую разницу между дисперсиями генеральной совокупности). Эта учеба разработал четыре процедуры, объединяющие существующие процедуры проверки равенства дисперсии и логики проверки эквивалентности.

Основываясь на частоте ошибок типа I и результатах мощности, тест эквивалентности Левена-Веллека-Уэлча, основанный на медиане, был наиболее надежной процедурой в тестируемых условиях с неизменно более высокой мощностью. над другими процедурами. Поэтому рекомендуется исследователям, желающим оценить равенство групповые отклонения.

Чтобы облегчить использование нашей недавно разработанной процедуры, функция для Levene-Wellek-Welch процедура, основанная на абсолютных отклонениях от медианы, была разработана в R (R Core Team, 2016) и доступен на http://cribbie.info.yorku.ca.

Ограничения

Хотя это исследование было сделано как можно более полным, существует много других условий. которые можно было бы протестировать для дальнейшей оценки новых процедур равенства дисперсий-эквивалентности. Трудно протестировать каждый сценарий данных, с которым может столкнуться исследователь. Однако результаты подтвердили цели этого исследования, в том, что фундаментальные недостатки традиционных тестов, основанных на различиях, были были выявлены, а недавно разработанные процедуры, основанные на эквивалентности, подвергались различным условиям данных для оценки их надежности. Кроме того, условия, выбранные для этого исследования, представляют собой общие условия анализа данных в педагогических и психологических исследованиях. Тем не менее, исследование производительности предложенных основанных на эквивалентности тестов однородности дисперсии в более широком диапазоне условий, безусловно, рекомендуется.

Более широкое ограничение процедур проверки эквивалентности в целом связано с принятием соответствующие интервалы эквивалентности. Определение интервала эквивалентности является наиболее сложной задачей. аспект проверки эквивалентности, потому что нет конкретных правил, которые помогли бы исследователям выбрать соответствующий интервал эквивалентности. Интервал эквивалентности должен быть выбран на основе знание своей области, их опыт работы с конструкциями и образцами, а также понимание того, как «бессмысленность» может быть определена количественно для их конкретного исследовательского вопроса. Хотя это может быть истолковано как ограничение, мы призываем исследователей тщательно обдумать бессмысленные различия между их группами при выборе интервалов эквивалентности вместо того, чтобы полагаться на эмпирические правила или общие рекомендации.

Прикладной пример

В этом разделе демонстрируется, как использовать критерий Левена-Уэллека Уэлча, наиболее эффективный критерий однородности дисперсии на основе эквивалентности с точки зрения мощности и контроля ошибок типа I, с использованием наглядный пример из психологических исследований. Мы также сравниваем результаты этого теста с результатами оригинальный медианный тест Левена с использованием тех же данных. Этот пример достигает двух целей: (1) продемонстрировать использование новой процедуры однородности дисперсии на основе эквивалентности; и (2) высветить фундаментальные недостатки исходных основанных на различиях типов Левена тестов на однородность отклонения.

Данные взяты из Arpin-Cribbie, Irvine, and Ritvo (2011). Участники, набравшие очень высокие баллы дезадаптивный перфекционизм случайным образом распределяли в одну из трех групп: отсутствие лечения, общий стресс управление или когнитивно-поведенческая терапия (КПТ). Участники оценивались по различным результатам на предварительном тесте и снова после вмешательства через 11 недель (посттест). Общий размер выборки составил 83. Представляло интерес обеспечить, чтобы три случайно распределенные группы не различались по исходным показателям с точки зрения центральной тенденции, а также гарантировать, что разброс баллов внутри каждой группы был одинаковым. сопоставимы между группами. В исходном исследовании рассматривалась эквивалентность групп по всем предварительным тестам. меры, но текущий пример проверяет только эквивалентность отклонений базовой меры Опросника познания перфекционизма (PCI; Flett, Hewitt, Blankstein, & Gray, 1998) с целью демонстрации. Дисперсия для группы управления стрессом (s2 = 110,79) и группы без лечения (s2 = 156,28) была одинаковой, но дисперсия для группы КПТ (s2 = 241,79) превышала два значения.
раз больше, чем в группе стресс-менеджмента.

Первоначальный тест Левена показал, что не было статистически значимых различий между групповые дисперсии при использовании популярных a Д .05, ФД 2,50, p Д .09. Тест Левена с использованием медианы (т. е. модификация Брауна-Форсайта теста Левена) также указывало на отсутствие статистически достоверные различия в групповой дисперсии, FMed D 2,10, p D 0,13. Затем был использован недавно разработанный тест эквивалентности Левена-Веллека-Уэлча, основанный на медиане, с установкой интервала эквивалентности е D 0,50. Этот Тест на эквивалентность показал, что дисперсии не были значительно эквивалентны (LWWmed D 6.42 > Fa, J-1,
Нью-Джерси, $\overline{п}$е2 Д 4.10). Таким образом, тесты, основанные на различиях, показали, что групповые дисперсии не отличались, но тест эквивалентности показал, что групповые дисперсии не были эквивалентны. Первопричиной этого произошло, обсуждалось во введении: поскольку размеры выборки в группах были относительно небольшой, способность обнаруживать даже нетривиальные различия в дисперсиях с помощью традиционных тестов была снижена. Следовательно, процедуры, основанные на различиях, объявили нетривиальные различия между групповыми дисперсиями эквивалентными, тогда как проверка эквивалентности показала, что разница в этих групповых дисперсиях превышает установленный предел эквивалентности. Использование новой процедуры, основанной на эквивалентности, гарантирует, что исследователи, которые оценивают равенство дисперсии, имеют действительный тест для оценки этой проблемы и будут, следовательно, сделать точные выводы о равенстве их групповых дисперсий.

Будущие направления

Будущие исследования должны включать обсуждение важности изучения вариаций, связанных с данными, и последствия однородности или неоднородности групповых дисперсий. Например, Bryk и Raudenbush (1988) предполагают, что неоднородность внутри групп может указывать на наличие
взаимодействие между характеристиками личности и принадлежностью к группе. Альтернативно, однородность групповые дисперсии при наличии средних различий могут указывать на то, что, даже если группы могут представляют разные популяции, они имеют сходство в составе, что может быть интересно исследовать. Однако дискуссии об однородности или неоднородности дисперсии с теоретической точки зрения не так популярны в науках об образовании и поведении, как в других дисциплинах. Например, Сагрестано, Хиви и Кристенсен (1998) утверждают, что различные точки зрения, как правило, сосредоточены на
различные аспекты изменчивости. Подход, основанный на индивидуальных различиях, фокусируется на межгрупповой изменчивости. пренебрегая внутригрупповой изменчивостью, в то время как социально-структурный подход фокусируется на внутригрупповой вариабельность, но может пренебречь межгрупповыми различиями. Будущие исследования могут быть сосредоточены на объединении этих

подходы, так что сравнение внутригрупповой изменчивости между группами становится важным исследовательским соображением, поэтому для этих целей исследования потребуется методологическая поддержка.

Выводы

Это исследование предоставило исследователям данные о проблемах с оценкой равенства дисперсий с помощью тестов, основанных на различиях. В частности, тесты на основе различий оценивают равенство дисперсий с неправильной точки зрения, побуждая исследователей поддерживать свои исследовательские гипотезы, не отвергая нулевую гипотезу. Таким образом, были предложены четыре новые процедуры, основанные на эквивалентности, для оценки равенства дисперсий. Из этих процедур критерий эквивалентности дисперсий Левена-Веллека-Уэлча, основанный на абсолютных отклонениях от медианы, был наиболее эффективной статистикой теста с точки зрения точной частоты ошибок типа I и наивысшей мощности для обнаружения эквивалентности в оцениваемых условиях. Следовательно, исследователи должны оценивать исследовательские гипотезы об эквивалентных дисперсиях населения, используя этот тест эквивалентности, основанный на медиане — Левена-Уэллека-Уэлча.

Примечание

1. Таблицы/рисунки полных результатов всех условий, оцененных в этом исследовании, можно получить, отправив электронное письмо соответствующему спонсирующий автор.
2. Ссылку на функцию R для вычисления теста Левена-Веллека-Уэлча можно найти по адресу http://cribbie.info.yorku.ca/.

использованная литература

Арпин-Крибби, К., Ирвин, Дж., и Ритво, П. (2011). Когнитивно-поведенческая терапия перфекционизма через Интернет: рандомизированное контролируемое исследование. Психотерапевтические исследования, 22, 194–207.

Боркенау, П., Гребчиков, М., Куппенс, П., Реало, А., и Аллик, Дж. (2013). Половые различия в изменчивости личности: исследование четырех образцов. Журнал личности, 81, 49–60.

Брэдли, СП (1978). Прочность? Британский журнал математической и статистической психологии, 81, 144–152.

Браун, М.Б., и Форсайт, А.Б. (1974). Надежные тесты на равенство дисперсий. Журнал Американской статистической ассоциации циация, 69, 364–367.

Брык, А.С., и Рауденбуш, С.В. (1988). Неоднородность отклонений в экспериментальных исследованиях: вызов общепринятым интерпретациям. Психологический бюллетень, 104, 396–404.

Коновер, В. Дж., Джонсон, М. Е., и Джонсон, М. М. (1981). Сравнительное исследование тестов на однородность дисперсий, с приложения к данным торгов на внешнем континентальном шельфе. Технометрика, 23, 351–361.

Крибби, Р.А., Груман, Дж.А., и Арпин-Крибби, Калифорния (2004). Рекомендации по применению тестов эквивалентности. Журнал клинической психологии, 60, 1–10.

Крибби, Р.А., Рагунанан, К., и Коунселл, А. (2016). Тестирование незначительного взаимодействия: последовательный и надежный подход. Br J Math Stat Psychol, 69, 159–174. doi: 10.1111/bmsp.12066.

Флетт, Г.Л., Хьюитт, П.Л., Бланкштейн, К.Р., и Грей, Л. (1998). Психологический дистресс и частота перфекционизма мышление. Журнал личности и социальной психологии, 75, 1363–1381.

Герцен, младший, и Крибби, Р.А. (2010). Обнаружение отсутствия ассоциации: подход к проверке эквивалентности. Британский журнал математической и статистической психологии, 63, 527–537.

Хартли, НО (1950). Использование диапазона в дисперсионном анализе. Биометрика, 37, 271–280.

Кесельман, Х. Дж., Игры, П. А., и Клинч, Дж. Дж. (1979). Тесты на однородность дисперсии. Коммуникации в статистике— Моделирование и вычисления, B8, 113–129.

Кесельман, Х. Дж., Хьюберти, С. Дж., Ликс, Л. М., Олейник, С., Крибби, Р., Донахью, Б., Ковальчук, Р. К., Лоуман, Л. Л., Петос Ки, М. Д., Кесельман, Дж. К., и Левин, младший (1998 г.).). Статистическая практика исследователей в области образования: анализ их ANOVA, MANOVA и ANCOVA. Обзор образовательных исследований, 68, 350–386.

Кесельман, Х.Дж., Уилкокс, Р.Р., Альгина, Дж., Отман, А.Р., и Фрадетт, К. (2008). Сравнительное исследование надежного теста на спред: стратегии асимметричной обрезки. Британский журнал математической и статистической психологии, 61, 235–253.

Левен, Х. (1960). Надежные тесты на равенство дисперсий. В I. Olkin, & et al. (ред.), Вклад в вероятность и статус тики, эссе в честь Гарольда Хотеллинга (стр. 278–292). Стэнфорд, Калифорния: Издательство Стэнфордского университета.

Лим, Т.С. и Ло, Вайоминг (1996). Сравнение тестов равенства дисперсий. Вычислительная статистика и анализ данных сис., 22, 287–301.

Нордстокке, Д. В., и Зумбо, Б. Д. (2010). Новый непараметрический критерий Левена для равных дисперсий. Psicologica, 31, 401–430.

Отман, А.Р., Кесельман, Х.Дж., Падманабхан, А.Р., Уилкокс, Р.Р., Альгина, Дж., и Фрадетт, К. (2004). Сравнение показателей «типичного» балла по группам лечения. Британский журнал математической и статистической психологии, 57, 215–234.

Парра-Фрутос, И. (2009). Поведение модифицированного теста Левена при ненормальном распределении данных. Вычисления Национальная статистика, 24, 671–693.

Основная команда R. (2016). R: Язык и среда для статистических вычислений. Вена, Австрия: R Foundation for Statis тические вычисления. Получено с https://www.R-project.org/

Робинсон, А.П., Дуурсма, Р.А., и Маршалл, Д.Д. (2005). Регрессионный тест эквивалентности для проверки модели: перенос бремени доказательства. Физиология деревьев, 25, 903–913.

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

Сагрестано, Л.М., Хиви, К.Л., и Кристенсен, А. (1998). Теоретические подходы к пониманию половых различий и сходства в конфликтном поведении. В DJ Canary и K. Dindia (Eds.), Половые различия и сходство в общении: Критические очерки и эмпирические исследования пола и гендера во взаимодействии (стр. 287–302). Махва, Нью-Джерси: Эрлбаум.

Сальгадо, Дж. Ф. (1995). Ситуационная специфичность и изменчивость валидности в пределах сеттинга. Журнал труда и организации Национальная психология, 68, 123–132.

Шуирманн, ди-джей (1987). Сравнение двух процедур односторонних тестов и силового подхода для оценки эквивалентность средней биодоступности. Журнал фармакокинетики и биофармацевтики, 15, 657–680.

Уэлч, Б.Л. (1951). О сравнении нескольких средних значений: Альтернативный подход. Биометрика, 38, 330–336.

Веллек, С. (2010). Проверка статистических гипотез эквивалентности (2-е издание). Бока-Ратон, Флорида: Chapman & Hall/CRC.

Уилкокс, Р.Р., Чарлин, В.Л., и Томпсон, К.Л. (1986). Новые результаты Монте-Карло по устойчивости дисперсионного анализа F, W, и F-статистика. Связь в статистическом моделировании и вычислениях, 15, 933–943.

Инь, Т.С. и Отман, А.Р. (2009). Когда t-критерий объединенных дисперсий терпит неудачу? Африканский журнал математики и коммуникаций Путер Научные исследования, 2, 56–62.

Циммерман, Д. В. (2004). Примечание о предварительных проверках равенства дисперсий. Британский журнал математики и статистики статистическая психология, 57, 173–181.