



## Тесты эквивалентности для сравнения коэффициентов корреляции и регрессии Алисса Коунселл\* и Роберт А. Крибби Йоркский университет, Торонто, Онтарио, Канада

Тесты эквивалентности являются альтернативой традиционным тестам, основанным на различиях, для демонстрации отсутствия связи между двумя переменными. Несмотря на то, что недавно было проведено несколько исследований, посвященных тестам эквивалентности для сравнения средних, было проведено мало исследований методов эквивалентности для оценки эквивалентности или подобия двух коэффициентов корреляции или двух коэффициентов регрессии. В текущем проекте предлагаются новые тесты для оценки эквивалентности двух коэффициентов регрессии или корреляции, полученных методом двух односторонних тестов (TOST) (Schuirmann, 1987, J. Pharmacokinet. Biopharm, 15, 657) и теста эквивалентности Андерсона и Хаука (1983, Stat. Commun., 12, 2663). Имитационное исследование использовалось для оценки эффективности этих тестов и сравнения их с общепринятым, но неподходящим методом оценки эквивалентности с использованием неотклонения нулевой гипотезы в тестах, основанных на различиях. Результаты показывают, что тесты на эквивалентность имеют более точную вероятность объявления эквивалентности, чем тесты на основе различий. Однако тесты на эквивалентность требуют больших размеров выборки, чтобы обеспечить достаточную мощность. Мы рекомендуем тест эквивалентности Андерсона-Хаука вместо метода TOST для сравнения коэффициентов корреляции или регрессии.

### 1. Введение

Исследователей часто интересует сравнение корреляции населения ( $\rho$ ) или коэффициентов регрессии ( $\beta$ ), и во многих случаях интерес заключается в демонстрации того, что коэффициенты эквивалентны. Например, Pillemer, Thomsen, Kuwabara и Ivcevic (2013) исследовали воспоминания о себе и о том, были ли эти воспоминания сосредоточены на достижениях или межличностных темах у участников из Дании и США. Одна из их гипотез заключалась в том, что «отношения между эмоциональной валентностью и тематическим содержанием должны быть одинаковыми в разных культурах» (Pillemer et al., 2013, стр. 213). Клогг, Петкова и Хариту (1995) описывают сравнение коэффициентов регрессии как наиболее распространенный метод в социальных исследованиях для сравнения двух объяснений конкретной переменной-предиктора. В некоторых случаях исследователи хотели бы продемонстрировать, что одна модель лучше другой, или просто что модели отличаются друг от друга. В других случаях исследователи хотели бы сделать вывод об эквивалентности наклонов регрессии двух групп. Paternoster, Brame, Mazerolle, and Piquero (1998) приводят пример, когда исследователи хотели бы определить, есть ли аналогичный эффект лечения от исправительной программы для впервые совершивших правонарушение по сравнению с повторными правонарушителями. Они обсуждают эту взаимосвязь в контексте традиционной проверки значимости нулевой гипотезы на основе различий (NHST), но представляют исследовательские вопросы, согласующиеся с

\* Корреспонденцию следует направлять Алиссе Коунселл, факультет психологии, Йоркский университет, Торонто, ON M3J 1P3, Канада (электронная почта: консультирование@yorku.ca).

## 2 Алисса Коунселл и Роберт А. Крибби

определение эквивалентности (например, равно ли  $b_1$   $b_2$ ?). Эти примеры освещают исследования, сравнивающие коэффициенты корреляции или регрессии независимых групп. Другие исследования включают сравнение коэффициентов корреляции, которые не являются независимыми. Например, исследователь может попытаться проверить новую шкалу депрессии, продемонстрировав, что корреляция между шкалой депрессии и показателем тревоги эквивалентна корреляции ранее утвержденной шкалы депрессии с тем же показателем тревоги (т. е.  $q_{12} = q_{13}$ ). ).

### 2. Современные подходы к тестированию эквивалентности

Проблема в психологии заключается в том, что исследователи поведения не имеют доступа к методам в статистическом программном обеспечении, которые оценивают эквивалентность коэффициентов корреляции или регрессии. В приведенных выше примерах не было легкодоступных тестов эквивалентности, которые исследователи могли бы использовать для надлежащей демонстрации эквивалентности между группами. Их стратегия (и наиболее распространенный метод в психологии) заключалась в проверке эквивалентности с использованием традиционной NHST, основанной на различиях, в соответствии с которой незначительная тестовая статистика считается индикатором эквивалентности. Такой подход ошибочен по двум основным причинам. Первая проблема носит теоретический характер: цель тестов, основанных на различиях, состоит в том, чтобы обнаружить различие или взаимосвязь, а не их отсутствие. Цитируя название статьи Алтмана и Блэнда (1995), «отсутствие доказательств не является свидетельством отсутствия». Другими словами, принятие нулевой гипотезы не подходит для установления эквивалентности, потому что исследователи никогда не смогут статистически определить, верна ли нулевая гипотеза. Вторая проблема имеет практическое значение для исследователей: использование тестов, основанных на различиях, с нулевой гипотезой об отсутствии различий имеет нежелательные статистические свойства, если целью является эквивалентность. Неопровержение традиционной нулевой гипотезы (например, отсутствие различий) по существу гарантировало бы эквивалентность при небольших размерах выборки из-за низкой способности находить различия. С другой стороны, было бы трудно сделать вывод об эквивалентности при больших размерах выборки, поскольку нулевая гипотеза почти всегда будет отвергнута из-за высокой мощности. Таким образом, статистически незначимые результаты могут быть следствием недостаточного размера выборки или плохого дизайна исследования. По нашему мнению, исследователи непреднамеренно используют неправильный анализ, поскольку еще не разработаны соответствующие методы или методики не получили широкого распространения в психологии. Ниже мы обсудим доступные методы проверки эквивалентности перед введением методов сравнения коэффициентов корреляции или регрессии.

### 3. Тестирование эквивалентности

Тесты эквивалентности разрабатывались и тестировались в течение последних нескольких десятилетий в области фармакокинетики, где исследователи часто хотят определить, эквивалентны ли эффекты двух лекарств. Доступны многочисленные методы, хотя некоторые из них используются более широко, чем другие (например, Anderson & Hauck, 1983; Brown, Hwang, & Munk, 1997; Ennis & Ennis, 2009; Hsu, Hwang, Liu, & Ruberg, 1994; Schuirmann, 1987; Вестлейк, 1972, 1976).

Первое упоминание о проверке эквивалентности в психологии было сделано Роджерсом, Ховардом и Весси (1993), которые обсуждали методы, описанные Шуирманном (1987) и Уэстлейком (1972, 1976). В их статье подчеркивается, как и почему психология должна использовать методы, основанные на эквивалентности, используемые в других дисциплинах. После статьи Роджерса и др. (1993) многие исследователи обсуждали полезность тестирования эквивалентности для психологических исследований (например, Крибби, Груман и Арпин-Крибби, 2004; Кендалл, Маррс-Гарсия, Нат и др.).

Шелдрик, 1999; Кертемон, 2011 г.; Моряк и Серлин, 1998). В частности, тесты на эквивалентность могут использоваться для ответа на основные исследовательские вопросы в психологии, а также для предоставления статистического анализа для обоснования других исследований или статистических решений. Например, тест на эквивалентность может обосновать объединение групп или установить, что две группы равны на исходном уровне до начала лечения (Rogers et al., 1993). Тестирование эквивалентности также может быть использовано в валидационных исследованиях. Чтобы проверить дискриминантную валидность, можно использовать тест эквивалентности на отсутствие ассоциации (например, Goertzen & Cribbie, 2010), чтобы продемонстрировать, что шкала не коррелирует с другой шкалой, измеряющей несвязанную конструкцию. Тесты эквивалентности также применяются для таких процедур, как метаанализ, когда данные исследований, а не групп, объединяются в соответствии с аналогичным критерием.

При проверке эквивалентности эквивалентность не определяется как строгая нулевая разность, потому что при ошибке выборки трудно найти математическую эквивалентность, даже если истинная разница между параметрами совокупности равна нулю. Эквивалентность подразумевает, что параметры достаточно похожи, так что нет никаких практических последствий в предположении, что они равны. Чтобы определить, являются ли параметры эквивалентными, исследователь должен выбрать интервал таким образом, чтобы различие параметров в выбранном интервале можно было считать несущественным. Этот интервал называется интервалом эквивалентности  $(d, d)$ , где  $d$  представляет собой расстояние от нуля до края интервала в любом направлении. Выбор подходящего интервала эквивалентности будет рассмотрен в разделе 9.4. Следуя структуре NHST, тесты эквивалентности используют нулевую и альтернативную гипотезы. Однако нулевая гипотеза состоит, например, в том, что существует нетривиальная разница между средними значениями совокупности, то есть разница лежит за пределами заранее заданного интервала эквивалентности. Альтернативная гипотеза состоит в том, что разница в средних попадает в интервал эквивалентности. Чтобы связать эти идеи с основной целью данной статьи, эквивалентность коэффициентов регрессии и корреляции включает в себя демонстрацию того, что разница между коэффициентами настолько мала, что любые различия можно считать тривиальными.

#### 4. Тесты эквивалентности, сравнивающие коэффициенты регрессии Мы

предлагаем два теста эквивалентности для сравнения независимых коэффициентов регрессии. Первый (TOST-b) представляет собой основанную на эквивалентности версию t-теста независимых коэффициентов регрессии, основанную на популярном методе двух односторонних тестов (TOST), изначально разработанном для оценки эквивалентности двух групповых средних (Schuirmann, 1987). ; Westlake, 1972). Первая нулевая гипотеза,  $H01: b_1 - b_2 \geq d$ , отвергается, если  $t_1 \leq -t(a, N/4)$  для нулевая гипотеза,  $H02: b_1 - b_2 \leq -d$ , отвергается, если  $t_2 \geq t(a, N/4)$ , где

$$\frac{b_1 - b_2 - d}{s_{b_1 - b_2}} \leq -t_{1-\alpha/2}(N/4); \quad \frac{b_1 - b_2 + d}{s_{b_1 - b_2}} \geq t_{1-\alpha/2}(N/4);$$

$b$  — коэффициент регрессии генеральной совокупности,  $b$  — коэффициент регрессии выборки,  $t$  распределен по  $N/4$  степеням свободы. Стандартную ошибку можно рассчитать по формуле

$$s_{b_1 - b_2} = \sqrt{\frac{s_{YX1}^2}{c_{X1}^2 n_1} + \frac{s_{YX2}^2}{c_{X2}^2 n_2}};$$

где  $s_{YX1}^2$  и  $s_{YX2}^2$  — дисперсии независимой переменной для каждой группы, а  $s_{YX1}^2$  и  $s_{YX2}^2$  — дисперсии ошибок для каждой группы. Отклонение как  $H01$ , так и  $H02$  необходимо

## 4 Алисса Коунселл и Роберт А. Крибби

сделать вывод, что разница коэффициентов регрессии находится в пределах интервала эквивалентности (т.е.  $d < b_1 \ b_2 < d$ ).

Наш второй тест эквивалентности коэффициентов регрессии (АН-b) основан на методе Андерсона. и тест эквивалентности Хаука (1983). Их процедура аппроксимирует нецентральный  $t$  распределение для определения  $p$ -значения для теста. Этот тест имеет только одну нулевую гипотезу,  $H_0: b_1 \ b_2 \ d$  или  $b_1 \ b_2 \ d$ , который отбрасывается при  $p \ a$ , где  $p$  аппроксимируется выражением

$$p \approx U \frac{|b_1 - b_2| d}{c_{b1 \ b2}} \cup \frac{|b_1 - b_2| d}{c_{b1 \ b2}} ;$$

в котором  $\Phi$  представляет собой стандартную нормальную функцию вероятности. Если  $p \ a$ , регрессия коэффициенты считаются эквивалентными.

## 5. Сравнение коэффициентов корреляции

## 5.1. Выборочное распределение коэффициента корреляции населения

На исследования с использованием коэффициентов корреляции влияет выборочное распределение параметр корреляции населения,  $\rho$ , потому что распределение становится все более искажается, когда  $\rho$  приближается к 1,00. При более высоких значениях  $\rho$  выборочное распределение не будет быть нормальным, и стандартную ошибку трудно оценить. Это имеет важное последствия для сравнения двух независимых коэффициентов корреляции при запуске  $t$ -критерий разницы (Howell, 2009). Фишер (1921) продемонстрировал, что путем преобразования  $r$  используя формулу

$$r_0 \approx \frac{1 + p}{1 - p} ;$$

$\rho$  станет приблизительно нормально распределенным вокруг  $\rho_0$  (преобразованного  $\rho$ ) с стандартную ошибку  $sr_0 = \frac{1}{N-3}$ . После преобразования  $r$  в  $r_0$ , проблема перекоса 1 выборочного распределения можно избежать, получив стандартизованную оценку разницы между независимыми  $\rho$ s. Поскольку значение  $d$  указано в линейной шкале,  $d$  должно быть рассчитывается таким образом, чтобы интервал находился в том же масштабе, что и преобразованный  $r$ s. Преобразование интервал эквивалентности оказался неэффективным, поскольку небольшие отклонения в разнице между  $r_0 \ 1 \ r_0 \ 2$  и  $d$  вызвало большие расхождения в тестовой статистике. Вместо этого  $d$  был изменен следующей статистикой, основанной на преобразовании  $r$  в  $z$  Фишера:

$$d_0 \approx \frac{1 + \frac{r_1 r_2}{2} \frac{A}{2}}{\left( 1 - \frac{r_1 r_2}{2} \frac{A}{2} \right)} \approx \frac{1 + \frac{r_1 r_2}{2} \frac{N_e}{2} \frac{A}{2}}{\left( 1 - \frac{r_1 r_2}{2} \frac{p}{2} \frac{A}{2} \right)} ;$$

Учитывая сложность преобразования  $r_1$  и  $r_2$ , мы рассмотрели несколько версий основанные на эквивалентности  $t$ -тесты, сравнивающие коэффициенты корреляции. Наш TOST- $q$  использовал преобразование выше, тогда как два других теста эквивалентности, сравнивающие корреляции (KTOST- $q$  и АН- $q$ ) вместо этого использовали модифицированную стандартную ошибку (описанную в следующем разделе).

## 5.2. Тесты эквивалентности, сравнивающие независимые коэффициенты корреляции

Первый предложенный тест для коэффициентов корреляции применил TOST к  $t$ -критерию преобразованные коэффициенты корреляции (ТОСТ- $q$ ). Подобно TOST, сравнивающему регрессию

коэффициентов, две нулевые гипотезы,  $H_{01}: \rho_1 = \rho_2$  и  $H_{02}: \rho_1 \neq \rho_2$ , отклоняются. при  $z_1 = z_2$  и  $z_2 = -z_1$ , где

$$z_1 = \frac{r_1 - r_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \quad z_2 = \frac{r_1 - r_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Отказ от обеих нулевых гипотез означает, что коэффициенты корреляции между две группы эквивалентны.

Вторым тестом, который мы исследовали (KTOST-q), был нетрансформированный TOST, предложенный Краатц (2007). Нулевая и альтернативная гипотезы аналогичны преобразованным тестам независимый кв. Однако она использовала следующую стандартную ошибку, описанную Олкиным и Финн (1995):

$$s_{r_1 - r_2} = \sqrt{\frac{\delta_1^2 r_1^2}{n_1 - 2} + \frac{\delta_2^2 r_2^2}{n_2 - 2}}$$

Тогда ее тестовая статистика

$$z_1 = \frac{r_1 - r_2}{s_{r_1 - r_2}}; \quad z_2 = \frac{r_1 - r_2}{s_{r_1 - r_2}}$$

где нулевые гипотезы отвергаются при  $z_1 = z_2$  и  $z_2 = -z_1$ , а отклонение обеих нулевая гипотеза означает, что коэффициенты корреляции эквивалентны между двумя группы.

Окончательный тест эквивалентности независимых коэффициентов корреляции, который мы предлагаем (AH-q) был основан на процедуре Андерсона и Хаука (1983). Мы использовали тот же стандарт ошибка, как Kraatz (2007), и после применения приближения Андерсона-Хаука, наша предложенная статистика была

$$p = \Phi \left( \frac{r_1 - r_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) \quad \text{или} \quad \Phi \left( \frac{r_1 - r_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right)$$

где  $\Phi$  представляет собой стандартную нормальную функцию вероятности. Когда  $p = \alpha$ , нуль гипотеза отвергается, и исследователь может сделать вывод, что коэффициенты корреляции эквивалентны.

### 5.3. Тесты эквивалентности, сравнивающие зависимые коэффициенты корреляции

Существует ряд основанных на различиях t-тестов, сравнивающих зависимые коэффициенты корреляции. в литературе (например, Dunn & Clark, 1969; Hotelling, 1931; Olkin, 1967; Williams, 1959).

Наши тесты эквивалентности для сравнения зависимых корреляций основаны на критериях Уильямса. модификация теста Хотеллинга для сравнения перекрывающихся зависимых корреляций. Этот тест был использован в качестве основы для предлагаемых тестов эквивалентности, потому что он был сравнен другим методом и был рекомендован за его общие статистические свойства.

(Boyer, Palachek, & Shucany, 1983; Hittner, May, & Silver, 2003; Steiger, 1980). Первый

предлагаемый тест эквивалентности, сравнивающий зависимые коэффициенты корреляции, основан на

6 Алисса Коунселл и Роберт А. Крибби

ТОСТ (ТОСТ-qD). Две его нулевые гипотезы, H01:  $q_{12} q_{13} = d$  и H02:  $q_{12} q_{13} = d$ , отвергаются при  $t_1 = t_1 a, N_3$  и  $t_2 = t_1 a, N_3$ , где

$$t_1 = \frac{1}{4} \frac{\delta r_{12} r_{13} \delta d b}{\delta N \frac{1}{N_3} \frac{\delta r_{12} r_{13} \delta d b}{j R_j b^2} P_1 p_{23} T^3};$$

$$t_2 = \frac{1}{4} \frac{\delta r_{12} r_{13} d b}{\delta N \frac{1}{N_3} \frac{\delta r_{12} r_{13} d b}{j R_j b^2} P_1 p_{23} T^3};$$

в котором  $j R_j \frac{1}{4} \delta r_{12} r_{13} \delta d b$  и  $\delta r_{12} r_{13} d b$  распределены по  $N_3$  степеням свободы. Когда обе t-статистики статистически значимы,  $r_{12}$  и  $r_{13}$  считаются эквивалентными, поскольку разница между корреляциями попадает в интервал эквивалентности.

Наш второй предложенный тест эквивалентности для зависимых коэффициентов корреляции (АН-qD) использует формулу Андерсона и Хаука (1983). При  $p = a$  нулевая гипотеза H0:  $q_{12} q_{13} = d$  или  $q_{12} q_{13} = d$  отвергается, и исследователь может сделать вывод об эквивалентности коэффициентов корреляции. Значение p тестов аппроксимируется как

$$p = \frac{1}{4} U \frac{\delta r_{12} r_{13} j d b}{\delta N \frac{1}{N_3} \frac{\delta r_{12} r_{13} j d b}{j R_j b^2} P_1 p_{23} T^3};$$

$$U \frac{\delta r_{12} r_{13} j d b}{\delta N \frac{1}{N_3} \frac{\delta r_{12} r_{13} j d b}{j R_j b^2} P_1 p_{23} T^3};$$

где  $\Phi$  представляет собой стандартную нормальную функцию вероятности.

## 6. Цели исследования.

Текущее исследование преследует четыре основные цели: продемонстрировать, что традиционные основанные на различиях тесты значимости нулевой гипотезы для сравнения коэффициентов корреляции или коэффициентов регрессии не подходят, когда цель состоит в том, чтобы определить, что эти коэффициенты эквивалентны; оценить и сравнить мощность и частоту ошибок типа I тестов эквивалентности с использованием диапазона размеров выборки, размеров эффекта и коэффициентов корреляции населения; дать рекомендации исследователям поведения в отношении того, какие из этих новых тестов наиболее практичны в использовании; и сделать доступным программное обеспечение с открытым исходным кодом для проведения рекомендуемого анализа (функции R для проведения процедур эквивалентности, обсуждаемых в этой статье, доступны по адресу <http://www.psych.yorku.ca/cribbie/rfunctions.html>).

## 7. Метод

Исследование методом Монте-Карло, состоящее из двух частей, использовалось для оценки частоты ошибок типа I и мощности вариаций тестов эквивалентности для сравнения коэффициентов регрессии и корреляции. В

часть первая (независимые  $q_s/b_s$ ), многомерные нормальные данные были смоделированы с использованием функции `rmvnorm` пакета `mvtnorm` в статистической программе с открытым исходным кодом R (R Development Core Team, 2010). В каждом условии значения для  $r_1$  и  $b_1$  были смоделированы из указанной корреляционной матрицы ( $q_1$ ), а  $r_2$  и  $b_2$  были смоделированы из второй корреляционной матрицы ( $q_2$ ). Во всех условиях первая матрица корреляции популяции оставалась неизменной, в то время как вторая матрица корреляции популяции отличалась от первой на указанную величину эффекта. Были использованы следующие статистические данные тестов эквивалентности: TOST независимых  $b_s$  (TOST-b), критерий Андерсона-Хаука для независимых  $b_s$  (АН-b), TOST независимых  $q_s$  (TOST-q), критерий Краатца независимых  $q_s$  (KTOST-q) и критерий независимых  $q_s$  Андерсона-Хаука (АН-q). Вторая часть (зависимые корреляции) включала одну матрицу корреляции населения с использованием того же метода генерации данных, который описан в первой части. Здесь корреляционная матрица отражала три популяционные корреляции для одной группы:  $q_{12}$ ,  $q_{13}$  и  $q_{23}$ . Во второй части были рассмотрены TOST зависимых  $q_s$  (TOST-qD) и тест Андерсона-Хаука зависимых  $q_s$  (АН-qD).

Управляемыми переменными исследования были размер выборки, размер эффекта и значения в матрицах корреляции населения (т. е.  $q_1$  и  $q_2$  для независимых корреляций и  $q_{12}$ ,  $q_{13}$  и  $q_{23}$  для зависимых корреляций). Интервал эквивалентности был установлен равным  $(.1, .1)$  для всех исследованных условий, поскольку  $.1$  является наименьшей корреляцией, признанной в порогах Коэна для малых, средних и больших корреляций (Cohen, 1962). Другими словами, разница в корреляциях на  $0,1$  считается минимально важной. Хотя определение  $d$  в этом исследовании было установлено несколько произвольно, поскольку не было заранее определенного исследовательского вопроса, следует отметить, что подходящее значение  $d$  будет варьироваться в зависимости от характера исследования. Величина эффекта, естественно, различалась для ошибки типа I и условий мощности. Для первой части (независимые группы) для тестов на эквивалентность было создано условие ошибки Типа I путем установки разницы между  $q_1$  и  $q_2$  точно равной  $d$ . Именно здесь можно было бы ожидать увидеть наибольшую вероятность ложных отклонений для тестов эквивалентности (т. е.,  $\alpha$ ), поскольку разница находится на границах  $d$ . В степенном условии для тестов эквивалентности параметры популяции были либо точно равными ( $q_1 = q_2 = 0$ ), либо различались на  $0,05$  ( $q_1 - q_2 = 0,05$ ). Для каждой повторности значения  $q_1$  и  $q_2$  варьировались от  $0,1$  до  $0,85$ . Были исследованы как условия

Условия моделирования были аналогичными для второй части (зависимые корреляции). Условие ошибки Типа I включало разницу между  $q_{12}$  и  $q_{13}$ , точно равную  $d$ , и двумя условиями мощности были  $q_{12} = q_{13}$  или  $q_{12} - q_{13} = 0,05$ . Для каждой повторности значения  $q_{12}$  и  $q_{13}$  варьировались от  $0,1$  до  $0,85$ , а значения  $q_{23}$  составляли  $0,2$ ,  $0,4$  и  $0,6$ . Мы выбрали эти значения параметров, чтобы отразить значения, с которыми обычно сталкиваются исследователи в области поведенческих наук.

Номинальный уровень  $\alpha$  был установлен на уровне  $0,05$  для всех анализов, и для каждого условия было проведено 5000 повторений. Более подробная информация об условиях моделирования и управляемых параметрах для обеих частей исследования моделирования представлена в таблице 1.

Традиционные тесты на основе различий, сравнивающие  $r_1$  и  $r_2$ ,  $b_1$  и  $b_2$  или  $r_{12}$  и  $r_{13}$ , также проводились для обеспечения сравнения между тестами на эквивалентность и методами, основанными на различиях, когда цель состоит в том, чтобы найти сходство между двумя переменными. Обратите внимание, что при сравнении тестов на эквивалентность и основанных на различиях терминология ошибки Типа I и мощности в этих двух методах меняются местами (например, условием ошибки Типа I для тестов эквивалентности является условие мощности в тестах на основе различий). Таким образом, количество отклонений ( $p < \alpha$ ) для тестов эквивалентности сравнивалось с количеством отказов ( $p > \alpha$ ) в методах, основанных на различиях, так что оба теста отражали вероятность заключения об эквивалентности. Брэдли (1978) предложил либеральный робастный интервал для частоты ошибок типа I:  $0,5\alpha$  и  $1,5\alpha$ . Таким образом, в текущем исследовании частота

Таблица 1. Сводка параметров моделирования

Параметр		Ценности
Д		.1
Независимые группы		
q1		.1, .2, .3, .4, .5, .6, .7, .8
d2		.1, .2, .3, .4, .5, .6, .7, .8, .15, 25, .35, .45, .55, .65, .75, .85
Величина эффекта	d2)	0 (мощность), 0,05 (мощность), 0,1 (ошибка типа I)
(q1 n (на группу)		25, 50, 75, 100, 125, 150, 250, 375, 500, 750, 1000, 1500
Зависимый		
q12		.1, .2, .3, .4, .5, .6, .7, .8
q13		.1, .2, .3, .4, .5, .6, .7, .8, .15, 25, .35, .45, .55, .65, .75, .85
q23		.2, .4, .6
Величина эффекта (q12	q13)	0 (мощность), 0,05 (мощность), 0,1 (ошибка типа I)
Н		50, 100, 250, 500, 1000

будет считаться приемлемым, если эмпирическая частота ошибок типа I находится в этих пределах. (0,025, 0,075).

8. Результаты

Учитывая, что картина результатов была одинаковой во многих условиях, только подмножество результаты представлены ниже. В частности, результаты для условий, где q1 q2 = 0,05, q12 q13 = 0,05, а результаты для условий неравного размера выборки опущены, поскольку они отражают результаты для q1 = q2, q12 = q13 и условия равного размера выборки.

8.1. Независимые коэффициенты корреляции и регрессии

8.1.1. Вероятность ложного заключения об эквивалентности

Вероятность ложного вывода об эквивалентности для каждого из тестов эквивалентности (p a) когда разница между qs или bs находится на уровне или за его пределами ( d; d ) и основанном на разнице испытаний (p > a) представлены в табл. 2. Результаты в целом оказались сходными для различных проверки эквивалентности независимых qs и bs, хотя некоторые важные отличия появился.

При меньших размерах выборки различия в частоте ошибок типа I наблюдались для ТОСТ-б и АН-б. В частности, частота ошибок типа I для ТОСТ-б была слишком консервативен при меньших размерах выборки; нулевая гипотеза никогда не отвергалась для выборки размеры ниже 250 на группу. Этого не было в случае с АН-б. Его частота ошибок типа I была стабильно около а независимо от размера выборки. При максимальном размере выборки частота ошибок для ТОСТ-б и АН-б были похожи. Такая же закономерность результатов наблюдалась для проверки эквивалентности независимых qs. Однако коэффициенты ошибок для всех корреляционно-ориентированных тесты на эквивалентность увеличивались по мере приближения q к 1,00. В нижнем диапазоне значений q результаты ТОСТ-q были аналогичны ТОСТ-б, но эмпирическая частота ошибок типа I увеличивалась по мере того, как значения q увеличились. Частота ошибок ТОСТ-q хорошо поддерживалась в группе for размеры выборки 100 для средних значений q, но были слишком либеральными для самых высоких значений из кв. Результаты КТОСТ-q были практически идентичны результатам ТОСТ-q и эмпирические коэффициенты ошибок типа I для АН-q были постоянно точными и минимально зависит от значения q.



Таблица 2. Вероятность ложного объявления эквивалентности для независимых групп

q1	q2 n (на группу)	Diff-b	Diff-q	TOST-b	АН-b	TOST-q	KTOST-q	АН-q
.1	.2	50	.912	0,918	0	0,047 0 ,048	0	0,052
		100	.894	0,894	0	0 ,048 0	0	0,045
		250	.799	.800	0		0	0,044
		500	.631	.630		0,050 0,048		0,049
		1000	.378	.369	0,046	,047	0,047	0,048
.2	.3	50	.914	0,912	0	0,052 0 ,049	0	0,043
		100	0,883	0,878	0	0	0	0,050
		250	.800	.785	0	0,053 0,047	0	0,047
		500	.631	.614	.003	,009 ,047 ,048	0,009	0,045
		1000	.364	.334	.046	,050 0,049 0,047	0,047	0,049
.3	.4	50	.924	.915	0	0,054 ,027 ,047	0	0,050
		100	0,882	.872	0	,045 ,050 0,054	0	0,052
		250	.779	.753	0	0,052 0,052 ,042	0	0,051
		500	.603	.556	0,008	,045 ,044 ,049	0,027	0,055
		1000	.322	.266	0,046	0,050	0,045	0,045
.4	.5	50	.910	.902	0		0	0,047
		100	.880	.859	0		0	0,049
		250	.756	.709	0			0,05
		500	.573	.493	0,022			0,041
		1000	.300	.198	0,044		0,041,044	0,044
.5	.6	50	.916	.895	0		0	0,050
		100	.862	.833	0			0,048
		250	0,728	.633	0		0,004	0,047
		500	.525	.363	0,032		,047	0,051
		1000	.243	.107	0,050		0,052	0,052
.6	.7	50	.906	0,872	0	0,048,004,048,047,050,054,054 0		0,051
		100	.850	0,775	0	0,051 0,051		0,052
		250	.688	0,505	0	,043 ,044 ,053	0,041	0,055
		500	.448	.209	0,039		0,053	0,052
		1000	.156	0,025	0,050	0,050 0,053	0,051	0,051
.7	.8	50	.894	.809	0	0,052 0	0	0,057
		100	.817	.638	0	.050 .020 ,044	0,016	0,056
		250	.612	.263	0,009	.055	0,052	0,052
		500	.340	0,052	0,052	.052 .064 ,050	.059	0,059
		1000	0,079	0,001	0,050	.056 ,049 ,072	.050	0,050
.8	.9	50	0,845	.557	0	.049 ,081 ,048	.055	0,049
		100	.727	.259	0	.078 ,050 ,079	.066	0,066
		250	.438	.014		.048 ,082	.057	0,057
		500	.142 0				.057	0,057
		1000	0,01	0	0,044,048		.059	0,059

Заметки. a = 0,05; 5000 повторений; Diff-b = регрессионный тест на основе различий, Diff-q = тест корреляции на основе различий, TOST-b = метод регрессионной эквивалентности двух односторонних тестов, АН-b = тест регрессионной эквивалентности Андерсона-Хаука, TOST-q = корреляция двух односторонних тестов метод эквивалентности, KTOST-q = метод корреляционной эквивалентности непреобразованных двух односторонних тестов Краатца, АН-q = критерий эквивалентности корреляции Андерсона-Хаука. Состояние ошибки I рода для тесты эквивалентности; условия питания для испытаний на основе разности. Цифры, выделенные жирным шрифтом, означают, что ставки попадают в либеральный робастный интервал Брэдли (0,025–0,075).

## 10 Алисса Коунселл и Роберт А. Крибби

Методы, основанные на различиях (Diff-q и Diff-b), показали несоответствующую скорость объявления эквивалентности, когда значения q для групповой совокупности не были равны. На результаты Diff-q больше повлияли значения q, чем Diff-b. Как видно из Таблицы 2, вероятность ложного объявления эквивалентности достигала 0,924 для размера выборки 50 на группу, при этом это число уменьшалось по мере увеличения размера выборки и увеличения значений q1 и q2. Для размеров выборки 1000 на группу и q5 до 0,7 или 0,8 показатели снизились до 0,01 или даже до нуля (что указывает на то, что тест теперь имеет достаточную мощность, чтобы обнаружить разницу q1).

$$q2 = 0,1).$$

## 8.1.2. Вероятность правильного заключения об

эквивалентности Результаты для вероятности заключения об эквивалентности представлены в таблице 3. При сочетании более высоких значений q с большими размерами выборки наблюдаются приемлемые уровни мощности (более 0,80) для всех тестов эквивалентности. Однако мощность теста TOST-b была нулевой для размеров выборки менее 250 на группу, когда значения q были меньше 0,5. AN-b показал низкую мощность при меньших размерах выборки, хотя в этих условиях он неизменно имел более высокую мощность, чем TOST-b. TOST-q дает результаты, аналогичные TOST-b, когда q составляет около 0,5. В более высоком диапазоне q с наибольшим размером выборки мощность TOST-q достигала 1,00. Аналогичные результаты наблюдались для KTOST-q. AN-q также демонстрировал проблемы с мощностью для меньших размеров выборки, особенно в более низком диапазоне q, но неизменно имел более высокую мощность, чем TOST-q и KTOST-q. При размерах выборки 1000 на группу мощность AN-q, TOST-q и KTOST-q была одинаковой.

Поскольку Diff-b и Diff-q предназначены для поиска различий, вероятность заключения об эквивалентности поддерживается примерно на уровне 100(1  $\alpha$ )% независимо от размера выборки, когда корреляции населения или коэффициенты регрессии точно равны. При наличии незначительных различий между группами (например, q1 q2 = 0,05) показатели окончательной эквивалентности ожидаемо уменьшались по мере увеличения размера выборки, поскольку способность традиционных тестов обнаруживать различия возрастает с увеличением N.

## 8.2. Зависимые коэффициенты корреляции 8.2.1.

Вероятность ложного вывода об эквивалентности

Показатели ложного вывода об эквивалентности при q23 = 0,4 представлены в таблице 4. Результаты были схожими для двух других проверенных значений q23, хотя меньшие значения привели к более консервативным показателям при небольших размерах выборки для TOST-qD. а более высокие значения привели к несколько более либеральной частоте ошибок. В целом для тестов зависимых корреляций выявлялась та же картина результатов, что и при сравнении q5 независимых групп.

В частности, TOST-qD показал частоту ошибок типа I, которая была слишком консервативной для размеров выборки менее 500 на группу, тогда как AN-qD продемонстрировала точную частоту ошибок для всех исследованных размеров выборки. Вероятность ложного объявления эквивалентности традиционным тестом (Diff-qD) была слишком высокой при малых размерах выборки и уменьшалась по мере увеличения размера выборки.

## 8.2.2. Вероятность правильного заключения об

эквивалентности Показатели заключения об эквивалентности при q23 = 0,4 представлены в таблице 5.

Уменьшение значения q23 продемонстрировало падение мощности как тестов эквивалентности, так и ее увеличение.

Таблица 3. Вероятность объявления эквивалентности для независимых групп

q1	q2 n (на группу)	Diff-b	Diff-q	TOST-b	АН-b	TOST-q	KTOST-q	АН-q	
.1	.1	50	.942	0,942 0	.946	0,057	0	0	0,056
		100	.948	0		0,061	0	0	0,067
		250	.950	0,952 0,948		0,085	0	0	0,095
		500	.949	.004 ,458		.181			0,183
		1000	0,950	0,950		.470	0,466	0,464	0,476
.2	.2	50	.951	0,951 0,949		0,053	0	0	0,058
		100	.948	0,956 0		0,065	0	0	0,071
		250	.956			.101	0	0	.101
		500	.953	0,952 0,008		.196	0,009	0,009	.202
		1000	.948	.947 ,476 ,950		.487	0,512	0,512	0,518
.3	.3	50	.949	0,948 0,949 0,943		.055	0	0	0,055
		100	.945	.021 ,950 ,517		.065	0	0	0,071
		250	.948	.956 0,952 0,956		.097	0	0	.103
		500	.945	0,950 ,059 ,954		.193	0,069	0,069	0,212
		1000	.949	.949 0,948 0,950		.524	0,573	0,573	0,575
.4	.4	50	.949	956 .138 .954		.062	0	0	0,059
		100	.953	.651 .945 0 .947 0		.073	0	0	0,069
		250	.951			.101	0	0	0,115
		500	.947			.210	0,188	0,187	0,268
		1000	.948			0,578	.702	0,699	.700
.5	.5	50	.946			0,059	0	0	0,061
		100	.947			0,070	0	0	0,081
		250	.951			.113	0		0,152
		500	.955			.252	0,359	0,356	0,383
		1000	.954			.653	0,821	0,820	0,820
.6	.6	50	.949			0,059	0	0	0,069
		100	0,953			0,075			0,105
		250	0,950	0,951 0,254		0,128	0,086	0,084	0,222
		500	0,953	0,951 .744		.313	0,593	0,588	0,591
		1000	.950	.954 0 .952		.744	0,935	0,934	0,934
.7	.7	50	.947			.054	0	0	0,082
		100	.945	0,953 0,947		.076	.005	0,004	0,129
		250	.949	.003 ,946 ,413		.176	.413	.403	0,427
		500	0,952	.947 ,859 ,945 ,948		.433	.846	.841	0,841
		1000	.950	0,947 ,167 ,943 0		.859	.993	0,993	0,933
.8	.8	50	.946	.969 ,949		.071	.023	018	0,137
		100	.949			.100	.265	0,242	0,316
		250	.941			0,266	.859	.843	0,843
		500	.947			0,672	.992	.991	0,991
		1000	.957			0,969	1.000	1.000	1.000

Заметки. a = 0,05; 5000 повторений; Diff-b = регрессионный тест на основе различий, Diff-q = тест корреляции на основе различий, TOST-b = метод регрессионной эквивалентности двух односторонних тестов, АН-b = тест регрессионной эквивалентности Андерсона-Хаука, TOST-q = корреляция двух односторонних тестов метод эквивалентности, KTOST-q = метод корреляционной эквивалентности непреобразованных двух односторонних тестов Краатца, АН-q = критерий эквивалентности корреляции Андерсона-Хаука. Состояние питания для тесты эквивалентности; Состояние ошибки типа I для тестов, основанных на различиях.

Таблица 4. Вероятность ложного объявления эквивалентности при  $q_{23} = 0,4$

q12	q13	H	Diff-qD	TOCT-qD	АН-qD
.1	.2	50	0,906	0	0,046
		100	.853	0	0,052
		250	.699	0	0,053
		500	.459	.045	0,051
		1000	.183	.060	0,061
.2	.3	50	.906	0	0,055
		100	.859	0	0,055
		250	.695	0	0,053
		500	.445	0,046	0,053
		1000	.170	0,055	0,055
.3	.4	50	.892	0	0,046
		100	.851	0	0,055
		250	.677	0	0,048
		500	.422	.046	0,051
		1000	.152	.054 0	0,055
.4	.5	50	.892		0,049
		100	.828	0	0,050
		250	.657	0	0,049
		500	.381	.047	0,050
		1000	.116	.053 0	0,053
.5	.6	50	0,887		0,053
		100	.810	0	0,052
		250	.611	.01	0,053
		500	0,319	.044	0,046
		1000	0,077	.048	0,048
.6	.7	50	0,872	0	0,055
		100	0,770	0	0,051
		250	.532	0,027	0,047
		500	.256	0,054	0,056
		1000	0,038	0,053	0,054
.7	.8	50	0,848	0	0,055
		100	.743	0	0,053
		250	.425	0,052	0,057
		500	0,135	0,051	0,052
		1000	0,011	0,052	0,052

Заметки.  $\alpha = 0,05$ ; 5000 повторений; Diff-qD = тест на основе различий, сравнивающий зависимые корреляции, TOST-qD = два односторонних теста, метод эквивалентности зависимой корреляции, АН-qD = Критерий эквивалентности Андерсона-Хаука зависимых коэффициентов корреляции. Состояние ошибки I рода для тесты эквивалентности; условия питания для теста на основе разности. Цифры, выделенные жирным шрифтом, означают, что ставки попадают в либеральный робастный интервал Брэдли (0,025–0,075).

привело к более высокой мощности как для TOST-qD, так и для АН-qD. Тот же образец результатов наблюдалось для разных значений  $q_{23}$ . Оба теста на эквивалентность продемонстрировали неадекватность мощность для размеров выборки менее 500. Однако преимущество по мощности наблюдалось для АН-qD при меньших размерах выборки с результатами, аналогичными TOST-qD при самой большой выборке. размер. Опять же, показатели заключительной эквивалентности для Diff-qD не отклонялись независимо от размера выборки, когда  $q_{12} q_{13} = 0$ , но уменьшается по мере увеличения размера выборки для  $q_{12} q_{13} = 0,05$  условие.

Таблица 5. Вероятность объявления эквивалентности при  $q_{23} = 0,4$ 

q12	q13	H	Diff-qD	TOCT-qD	АН-qD
.1	.1	50	0,948	0	0,060
		100	.952	0	0,078
		250	.945		0,136
		500	.949	0,316	.360
		1000	0,951	,786	0,788
.2	.2	50	0,957	0	0,059
		100	.950	0	0,073
		250	.946	0	.141
		500	.947	.308	0,347
		1000	.947	0,782	0,783
.3	.3	50	.953	0	0,062
		100	.948	0	0,076
		250	.951	0	0,146
		500	.953	.356	.390
		1000	.948	.812	0,813
.4	.4	50	.950	0	0,062
		100	.954	0	0,084
		250	.949	.003	0,166
		500	.951	.391	.415
		1000	.947	.833	0,835
.5	.5	50	.961	0	0,071
		100	.954		0,088
		250	.952	0,024	.194
		500	0,953	,492	0,507
		1000	0,955	0,895	0,895
.6	.6	50	.951	0	0,064
		100	.949	0	0,090
		250	0,952	0,092	0,228
		500	.944	.582	0,585
		1000	.949	.941	0,942
.7	.7	50	.949	0	0,074
		100	.954	0	0,117
		250	.957	0,241	.312
		500	.948	0,745	0,747
		1000	.949	0,979	0,980
.8	.8	50	.946	002	0,088
		100	.950	0,024	0,164
		250	.945	.536	0,551
		500	.948	.914	0,915
		1000	.949	.999	0,999

Заметки.  $\alpha = 0,05$ ; 5000 повторений; Diff-qD = тест на основе различий, сравнивающий зависимые корреляции, TOST-qD = два односторонних теста, метод эквивалентности зависимой корреляции, АН-qD =

Критерий эквивалентности Андерсона-Хаука зависимых коэффициентов корреляции. Состояние питания для тесты эквивалентности; Состояние ошибки типа I для тестов, основанных на различиях.

## 9. Обсуждение

Тестирование эквивалентности имеет много полезных применений в психологии. Однако немногие исследователи поведения используют тесты на эквивалентность, несмотря на то, что цели исследования совпадают с поиском эквивалентности. Одной из областей, где тесты на эквивалентность для сравнения независимых коэффициентов регрессии или корреляции особенно актуальны, являются исследования, посвященные культурному сходству. Исследователи часто стремятся продемонстрировать, что отношения между переменными (например, социальная поддержка и депрессия) согласуются между странами или культурами. Несколько примеров других дизайнов исследований в психологии, которые выиграют от проверки эквивалентности, представлены в предыдущих статьях (например, Cribbie et al., 2004; Kendall et al., 1999; Rogers et al., 1993; Stegner, Bostrom, & Greenfield, 1996).

Это исследование предоставляет исследователям подходящие методы для ответа на свои исследовательские вопросы, когда интерес заключается в демонстрации эквивалентности или сходства коэффициентов корреляции или регрессии.

### 9.1. Неуместность тестов, основанных на различиях. Поскольку

методы проверки эквивалентности редко используются в психологии, исследователи поведения продолжают использовать методы, основанные на различиях, при которых неотвержение нулевой гипотезы интерпретируется как эквивалентность. При использовании для демонстрации сходных отношений традиционные тесты значимости нулевой гипотезы, основанные на различиях, демонстрируют неадекватные показатели заключительной эквивалентности. Например, вероятность обнаружения эффекта, связанного с альтернативной гипотезой, должна быть напрямую связана с размером выборки — увеличение размера выборки должно увеличить шансы обнаружения эффекта. Поскольку традиционные методы, основанные на различиях, предназначены для проверки различий, поиск эквивалентности противоречит их цели. Это приводит к обратной связи с размером выборки и обнаружению эффекта, согласующегося с гипотезой исследователя. При меньших размерах выборки эквивалентность устанавливается приблизительно в 100(1 $\alpha$ ) % случаев с использованием методов, основанных на различиях. Таким образом, когда значения  $q$  точно эквивалентны, способность найти эквивалентность не изменилась по мере увеличения размера выборки. Когда были небольшие различия между  $q$ s (но все еще в пределах интервала эквивалентности), вероятность объявления эквивалентности была высокой при небольших размерах выборки, но уменьшалась по мере увеличения размера выборки.

Наблюдение за уменьшением способности найти интересующий исследователя эффект при увеличении размера выборки противоречит исследовательской практике. Исследователи стремятся к большому размеру выборки, чтобы обеспечить достаточную мощность и обобщаемость результатов. Эта проблема предполагает, что методы, основанные на различиях, непригодны для проверки эквивалентности. Вместо этого исследователи должны использовать тесты эквивалентности, если они хотят найти эквивалентность двух коэффициентов корреляции или регрессии. Эти тесты теоретически обоснованы и обладают желательными статистическими свойствами (например, способность найти интересующий эффект увеличивается с увеличением размера выборки; нулевые гипотезы отвергаются, а не принимаются).

### 9.2. Различия между тестами эквивалентности Как было

отмечено в литературе, процедура TOST может быть слишком консервативной, когда мощность мала или увеличивается дисперсия (Berger & Hsu, 1996; Brown et al., 1997). Это исследование моделирования подтвердило этот вывод для TOST-b, TOST-q, KTOST-q и TOST-qD.

Напротив, тесты эквивалентности на основе АН имели точную частоту ошибок типа I для всех размеров выборки и значений  $q$ . Эти результаты противоречат литературным данным, в которых сообщается, что частота ошибок процедуры Андерсона-Хаука может быть чрезмерно либеральной (Brown et al.,

1997 год; Эннис и Эннис, 2009 г. ; Нам и Мунк, 1994). Мощность каждого из тестов эквивалентности была почти одинаковой для больших размеров выборки. В условиях меньшего размера выборки тесты Андерсона-Хаука неизменно имели более высокую мощность, чем тесты TOST, хотя они все еще были неудовлетворительными по стандартам исследователей поведения.

### 9.3. Низкая мощность

Как уже упоминалось, исследователи в идеале стремятся к мощности не менее 0,80. Глядя на результаты мощности тестов эквивалентности, ясно видно, что мощность неадекватна для размеров выборки менее 1000 на группу для независимых групп или 500 в схеме зависимой корреляции. Мощность всех тестов эквивалентности считается низкой для типов размеров выборки, которые обычно встречаются в психологии. Однако современные t-тесты, сравнивающие коэффициенты регрессии или корреляции, демонстрируют меньшую мощность по сравнению с другими популярными статистическими анализами, такими как анализ разностей средних (Howell, 2009). Поскольку  $q$  должно находиться в диапазоне от 1,00 до +1,00, разница между двумя  $q$  относительно их стандартной ошибки обычно будет ниже по сравнению со строгой средней разницей без ограниченной шкалы. Другими словами, основанные на различиях тесты значимости нулевой гипотезы, сравнивающие регрессию или корреляции, также демонстрируют низкую мощность, когда используются для их надлежащей цели поиска различий. Исследователи, целью которых является поиск эквивалентности или различия между коэффициентами регрессии/корреляции, должны знать, что этот план исследования требует больших размеров выборки для обеспечения достаточной мощности. Еще один момент, о котором стоит упомянуть, заключается в том, что мощность также связана с размером интервала эквивалентности.

Если указанный интервал эквивалентности достаточно строгий или мал, мощность будет ниже, чем если бы использовался больший интервал эквивалентности. Учитывая, что в исследовании использовалось моделирование и не было включено существенной гипотезы относительно эквивалентности коэффициентов регрессии или корреляции, возможно, что использованный нами интервал эквивалентности мог привести к меньшей мощности, чем могли бы наблюдать прикладные исследователи.

### 9.4. Выбор интервала эквивалентности Величина

интервала эквивалентности является важной темой при обсуждении тестов эквивалентности. Наш интервал эквивалентности  $d$  был установлен равным 0,1. Мы выбрали это значение, потому что решили, что оно представляет наименьшую значимую разницу между двумя коэффициентами корреляции или регрессии в контексте нашего исследования моделирования. Значения, превышающие 0,1, могут иметь важное значение для отношения между двумя коэффициентами корреляции. Меньшее значение  $d$ , например 0,05, может быть слишком строгим интервалом для объявления эквивалентности, особенно с учетом того, что t-тесты, сравнивающие коэффициенты регрессии или корреляции, имеют меньшую мощность, чем другие виды анализа, такие как t-тесты, сравнивающие средние значения.

Стандарты Козна для величины эффекта также считают .1 небольшой размер эффекта среди корреляционных различий (Cohen, 1962). Хотя эти стандарты могут служить полезным ориентиром при выборе интервала эквивалентности, наиболее важным аспектом при выборе  $d$  является априорный выбор значения, которое можно считать наибольшей разницей между коэффициентами, которую исследователь счел бы несущественной. Интервал эквивалентности может принимать форму конкретных значений, как это было сделано в моделирующем исследовании, или разница может соответствовать конкретной стандартизированной величине эффекта (например,  $r$ , равное 0,01, соответствует 1% общей дисперсии, объясненной  $M$ ). Чотуравоватфаникая предвзятость, поскольку исследователь может выбрать любое значение, которое ему нравится для  $d$ , это не является обоснованным опасением, поскольку  $d$  должно быть теоретически обосновано. Роджерс и др. (1993, стр. 564) мудро заметил, что «как и в случае любого статистического анализа, процедуры эквивалентности должны включать вдумчивое планирование со стороны

следователь». Пока исследователь выбирает значение  $d$  перед сбором данных, и это значение подходит для решаемой исследовательской задачи, исследователь никоим образом не искажает свои результаты.

### 9.5. Рекомендации

Основываясь на результатах текущего исследования моделирования, традиционные  $t$ -тесты на основе различий для коэффициентов регрессии и корреляции не рекомендуются, когда целью исследователя является установление эквивалентности. Частота ложных выводов об эквивалентности у них слишком высока, и лучший способ найти эквивалентность — это иметь небольшой размер выборки. Тридцать лет назад Блэквелдер (1982, стр. 346) лаконично заявил, что « $r$  является мерой доказательств против нулевой гипотезы, а не в ее пользу, и недостаточные доказательства для отклонения нулевой гипотезы не предполагают достаточных доказательств для ее принятия». Многие утверждают, что исследователи никогда не должны принимать нулевую гипотезу, но исследователи продолжают делать это, когда используют несущественные результаты для оправдания равенства.

В заключение, NHST, основанный на различиях, никогда не будет действительным статистическим методом, если цель исследователя состоит в том, чтобы продемонстрировать эквивалентность. Процедуры проверки эквивалентности являются подходящими методами для нахождения эквивалентности параметров генеральной совокупности, таких как  $q$  или  $b$ . В частности, мы рекомендуем тест эквивалентности Андерсона-Хаука для сравнения коэффициентов регрессии или корреляции. Было обнаружено, что эта процедура поддерживает точную частоту ошибок типа I и продемонстрировала более высокую мощность, чем TOST, при меньших размерах выборки. Учитывая меньшую мощность статистических тестов, сравнивающих коэффициенты регрессии и корреляции, также рекомендуется, чтобы исследователи собирали большой объем данных перед выполнением этих статистических процедур, чтобы обеспечить адекватную мощность своих гипотез.

### 9.6. Ограничения и будущие направления

Одним из потенциальных ограничений является то, что все используемые данные были смоделированы. Здесь все данные были нормально распределены, тогда как данные, обычно встречающиеся в психологии, часто демонстрируют некоторую асимметрию и эксцесс. Важная задача тогда состояла бы в том, чтобы изучить выполнение тестов эквивалентности, используя данные, более типичные для тех, которые встречаются в психологических исследованиях. Кроме того, можно было проверить и другие условия, такие как другие значения  $d$  или использование других размеров выборки. Тем не менее, эти результаты, вероятно, можно было бы предсказать на основе результатов текущего исследования и, вероятно, они не нужны. Одним из будущих направлений является изучение методов повышения мощности тестов эквивалентности при сравнении коэффициентов регрессии и корреляции для их использования в психологии, учитывая их низкую мощность.

#### использованная литература

- Альтман, Д. Г., и Бланд, Дж. М. (1995). Отсутствие доказательств не является доказательством отсутствия. Британский медицинский Журнал, 311, 485. doi:10.1136/bmj.311.7003.485
- Андерсон, С., и Хаук, В.В. (1983). Новая процедура проверки эквивалентности в сравнительных исследованиях биодоступности и других клинических исследованиях. Статистика и коммуникации - теория и методы, 12, 2663–2692. doi: 10.1080/03610928308828634
- Бергер, Р.Л., и Ской, Дж.К. (1996). Испытания на биоэквивалентность, тесты пересечения и доверительные наборы эквивалентности. Статистическая наука, 11, 283–319. doi: 10.1214/cc/1032280304



- Блэквелдер, WC (1982). «Доказательство нулевой гипотезы» в клинических испытаниях. контролируемые клинические испытания, 3, 345–353. doi: 10.1016/0197-2456(82)90024-1
- Бойер, Дж. Э., Палачек, А. Д., и Шукани, В. Р. (1983). Эмпирическое исследование связанных коэффициентов корреляции. Журнал образовательной и поведенческой статистики, 8, 75–86. doi: 10.3102/10769986008001075
- Брэдли, СП (1978). Прочность? Британский журнал математической и статистической психологии, 31, 144–152. doi:10.1111/j.2044-8317.1978.tb00581.x
- Браун, Л.Д., Хван, Дж.Т.Г., и Мунк, А. (1997). Беспристрастный тест на проблему биоэквивалентности. Анналы статистики, 25, 2345–2367. doi: 10.1214/aos/1030741076
- Клогг, К., Петкова, Э., и Хариту, А. (1995). Статистические методы сравнения коэффициентов регрессии между моделями. Американский журнал социологии, 100, 1261–1293. doi: 10.1086/230638
- Коэн, Дж. (1962). Статистическая сила аномально-социальных психологических исследований: обзор. Журнал аномальной и социальной психологии, 65, 145–153. doi: 10.1037/h0045186
- Крибби, Р.А., Груман, Дж.А., и Арпин-Крибби, Калифорния (2004). Рекомендации по применению тестов эквивалентности. Журнал клинической психологии, 60, 1–10. doi:10.1002/jclp.10217 Данн, О.Дж., и Кларк, В. (1969). Коэффициенты корреляции, измеренные на одних и тех же людях. Журнал Американской статистической ассоциации, 64, 366–377. doi: 10.1080/01621459.1969.10500981 Эннис, Д.М., и Эннис, Дж.М. (2009). Проверка гипотез на эквивалентность, заданную на симметричных открытых интервалах. Коммуникации в статистике - теория и метод, 38, 1792–1803. doi: 10.1080/03610920802460787 Фишер, Р.А. (1921). О вероятной ошибке коэффициента корреляции, полученного по малой выборке. Метрон, 1, 3–32. Получено с <http://digital.library.adelaide.edu.au/dspace/bitstream/2440/15169/1/14.pdf>
- Герцен, младший, и Крибби, Р.А. (2010). Обнаружение отсутствия ассоциации: подход к проверке эквивалентности. Британский журнал математической и статистической психологии, 63, 527–537. doi: 10.1348/000711009C475853
- Хиттнер, Дж. Б., Мэй, К., и Сильвер, Северная Каролина (2003). Оценка методом Монте-Карло тестов для сравнения зависимых корреляций. Журнал общей психологии, 130, 149–168. doi: 10.1080/00221300309601282
- Хотеллинг, Х. (1931). Обобщение коэффициента Стьюдента. Анналы математической статистики, 2, 360–378. doi: 10.1214/aoms/1177732979 Хауэлл, округ Колумбия (2009). Статистические методы для психологии (7-е изд.). Белмонт, Калифорния: Томпсон.
- Хсу, Дж. К., Хван, Дж. Т. Г., Лю, Х., и Руберг, С. Дж. (1994). Доверительные интервалы, связанные с тестами на биоэквивалентность. Биометрика, 81, 103–114. doi: 10.1093/биомет/81.1.103
- Кендалл, ПК, Маррс-Гарсия, А., Нэт, С.Р., и Шелдрик, Р.К. (1999). Нормативные сравнения для оценки клинической значимости. Журнал консалтинга и клинической психологии, 3, 285–299. doi: 10.1037/0022-006X.67.3.285
- Краатц, М. (2007). Проверка корреляционной эквивалентности. Неопубликованная магистерская диссертация, Университет Вандербильта, Нэшвилл, Теннесси.
- Нам, Ю.В., и Мунк, А. (1994). О методе комбинирования двойного t-критерия и критерия Андерсона-Хаука. Биометрия, 50, 884–886.
- Олкин, И. (1967). Корреляции пересмотрены. В J.C. Stanley (Ed.), Улучшение экспериментального дизайна и статистического анализа (стр. 102–128). Чикаго: Рэнд МакНалли.
- Олкин, И., и Финн, Д.Д. (1995). Редукция корреляций. Психологический бюллетень, 118, 155–164. doi: 10.1037/0033-2909.118.1.155 Патерностер Р., Бrame Р., Мазероль П. и Пикеро А. (1998). Использование корректного статистического теста на равенство коэффициентов регрессии. Криминология, 36, 859–866. doi: 10.1111/j.1745-9125.1998.tb01268.x
- Пилемер, Д.Б., Томсен, Д., Кувабара, К.Дж., и Ивчевич, З. (2013). Чувство хорошего и плохого по поводу прошлого и будущего себя. Память, 2, 210–218. doi: 10.1080/09658211.2012.720263 Quertemont, E. (2011). Как статистически показать отсутствие эффекта. Psychologica Belgica, 51 год, 109–127. doi: 10.5334/pb-51-2-109

18 Алисса Коунселл и Роберт А. Крибби

Основная группа разработки R (2010 г.). R: Язык и среда для статистических вычислений.

Вена: R Foundation for Statistical Computing [руководство по компьютерному программному обеспечению].

Получено с <http://www.R-project.org> 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. doi: 10.1037/0033-2909.113.3.553

Шуирманн, ди-джей (1987). Сравнение двух процедур односторонних тестов и мощностного подхода для оценки эквивалентности средней биодоступности. Журнал фармакокинетики и биофармацевтики, 15, 657–680. doi: 10.1007/BF01068419

Моряк, Массачусетс, и Серлин, Р.К. (1998). Доверительные интервалы эквивалентности для двухгрупповых сравнений средних. Психологические методы, 3, 403–411. doi:10.1037/1082-989X.3.4.403 Стегнер, Б.Л., Бостром, А.Г., и Гринфилд, Т.К. (1996). Тестирование эквивалентности для использования в психосоциальных исследованиях и исследованиях услуг: введение с примерами. Образование и планирование программ, 19, 193–198. doi: 10.1016/0149-7189(96)00011-0 Steiger, JH (1980). Тесты для сравнения элементов корреляционной матрицы. Психологический бюллетень, 87, 245–251. doi: 10.1037/0033-2909.87.2.245 Westlake, WJ (1972). Использование доверительных интервалов в анализе сравнительных исследований биодоступности.

Журнал фармацевтических наук, 61, 1340–1341. doi: 10.1002/jps.2600610845 Westlake, WJ (1976).

Симметричные доверительные интервалы для исследований биоэквивалентности. Биометрика, 32, 741–744. Получено с <http://www.ncbi.nlm.nih.gov/pubmed/1009222>.

Уильямс, Э. Дж. (1959). Сравнение регрессионных переменных. Журнал Королевского статистического общества, серия B, 21, 396–399. Получено с <http://www.jstor.org/stable/2983809>.

Поступила в редакцию 24 февраля 2014 г.; исправленная версия получена 26 июля 2014 г.