Journal of Statistical Theory and Practice (2024) 18:36
https://doi.org/10.1007/s42519-024-00382-7

ORIGINAL ARTICLE

The Extended Anderson and Hauck Tests and Sample Size Procedures for Equivalence Assessment in Simple Linear Regressions

Gwowen Shieh[1]

Accepted: 12 May 2024 / Published online: 26 June 2024

Abstract
This study describes extended Anderson and Hauck procedures for equivalence testing of slope coefficients and mean responses in one and two regression lines. The general formulation of asymmetric equivalence ranges permits a wide variety of equivalence questions to be tested for a target magnitude or a negligible value. Specifically, the equivalence tests are useful for assessing negligible trend and similar response in a single regression line, and for evaluating unimportant interaction-moderation effect and comparable simple effect between two linear regression lines. The associated power functions and sample size procedures are also derived and compared under the random and fixed model settings. According to the analytic justification and empirical assessment, the exact approaches have a clear advantage over the approximate formulas for accommodating the full stochastic nature of both the response and predictor variables. Computer algorithms are also provided for conducting the proposed equivalence tests, power calculations, and sample size determinations in simple linear regressions.

Keywords ANCOVA · TOST · Moderation · Power · Sample size

1 Introduction

Many studies are designed explicitly to show that there is an absence of effects of competing scenarios or theories. However, they sometimes base their findings on failing to reject a null hypothesis rather than confirming a hypothesis of equivalence. For comparison of treatment effects, the traditional hypothesis test of difference aims to determine whether the treatment effects differ from one another. Under such condition, the traditional difference tests are inappropriate to establish equivalence, because

✉ Gwowen Shieh
gwshieh@nycu.edu.tw

[1] Department of Management Science, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan

failing to reject a no-difference hypothesis test does not necessarily support the conclusion of equivalence. There has been a growing awareness and demand of appropriate techniques for assessing equivalence and similarity in the behavioral and managerial literature. For example, related discussions of theoretical perspectives and practical issues can be found in Cashen and Geifer [1], Cortina and Folger [2], Edward and Berry [3], Frick [4], Rogers, Howard, and Vessey [5], Seaman and Serlin [6], Stanton [7], Stegner, Bostrom, and Greenfield [8], and Steiger [9], among others.

To assess an observed effect size that is clinically negligible or practically non-important, the recommended equivalence test is to ascertain whether the observed effect size falls inside the selected equivalence range. The technical discussion and fundamental review of different types of mean equivalence tests were presented in Berger and Hsu [10], Meyners [11], and Schuirmann [12]. Despite there are more powerful tests, two prominent procedures have received considerable attention in the literature. They are the two one-sided tests (TOST) method of Schuirmann [13] and Westlake [14] and the equivalence approach of Anderson and Hauck [15] and Hauck and Anderson [16]. These two procedures of mean equivalence admit a simple methodological reform for assessing equivalence. Their flexible settings allow generalizations to more complex experimental designs. Accordingly, Dixon and Pechmann [17], and Schmidt and Meyer [18] have extended the TOST to assess whether the linear trend is practically negligible in linear regressions. Also, Counsell and Cribbie [19] described an extension of the Anderson and Hauck procedure for comparing the slope coefficients of two regression lines.

Despite the conservative nature, TOST maintains a good control of Type I error rate at the specified level. However, the actual Type error rate of TOST can be substantially less than the nominal level and the rejection region can be empty when the equivalence ranges are narrow, particularly with small sample sizes. Across the practical and diverse research designs for equivalence assessment, the undertaken equivalence bounds and associated sample sizes may not be all that large. Under such circumstances, it is of methodological concern to consider alternative procedures with proper rejection region and good Type I error control. On the other hand, the normal approximation presented in Counsell and Cribbie [19] for p-value calculations is only one of the three possible methods proposed in Anderson and Hauck [15]. Following the results of an extensive simulation study, Anderson and Hauck [15] recommended the central-t approach, instead of the least accurate normal approximation. In view of the absence of vital clarification for theory development and supportive technique, it is desirable to properly generalize the Anderson and Hauck procedure for linear regression analysis.

The present article aims to contribute to the development of equivalence methodology for linear regressions in three aspects. First, using the central-t approximation, extended Anderson and Hauck procedures are presented for equivalence testing of slope coefficient and mean response in one and two regression lines. The general formulation of asymmetric equivalence ranges permits a wide range of equivalence questions to be tested. Consequently, they are useful for assessing negligible trend and similar response in a single regression line, and for evaluating unimportant interaction-moderation effect and comparable simple effect between two linear regression lines. Second, the associated power functions and sample size procedures are also derived and compared under the random and fixed model settings. According to the analytic

justification and empirical assessment, the exact approaches have a clear advantage over the approximate formulas for accommodating the full stochastic nature of both the response and predictor variables. It should be noted that exact power and sample size calculations were not addressed in Counsell and Cribbie [19]. Third, the proposed equivalence techniques are not available in popular software packages. Computer algorithms are provided for critical value computations, power calculations, and sample size determinations of the extended Anderson and Hauck procedures. The suggested power and sample size calculations should be useful for planning equivalence studies about the much-discussed appraisals of interaction-moderation effect and simple effect in behavioral and management research.

2 Single Regression Line

The simple linear regression model is of the form

$$Y_i = \beta_0 + X_i\beta_1 + \varepsilon_i, \tag{1}$$

where Y_i is the response score of the ith subject, β_0 is the intercept, β_1 is the slope coefficient, X_i is the predictor score of the ith subject, and ε_i are $iid\ N(0, \sigma^2)$ random variables, $i = 1, \ldots, N$. The least squares estimator $\hat{\beta}_1$ of slope coefficient β_1 has the following distribution

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2/SSX), \tag{2}$$

where $SSX = \sum_{i=1}^{N} (X_i - \bar{X})^2$ and $\bar{X} = \sum_{i=1}^{N} X_i/N$. Also, $\hat{\sigma}^2 = SSE/\nu$ is the usual unbiased estimator of σ^2 where SSE is the error sum of squares and $\nu = N - 2$. Moreover, $V = SSE/\sigma^2 \sim \chi^2(\nu)$, where $\chi^2(\nu)$ are chi-square distribution with ν degrees of freedom.

To detect the difference of slope coefficient in terms of H_0: $\beta_1 = \beta_{10}$ versus H_1: $\beta_1 \neq \beta_{10}$, the test statistic has the form

$$T_{S0} = \frac{\hat{\beta}_1 - \beta_{10}}{(\hat{\sigma}^2/SSX)^{1/2}} \tag{3}$$

The null hypothesis is rejected at the significance level α if

$$|T_{S0}| > t_{\nu,\alpha/2} \tag{4}$$

where $t_{\nu,\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of $t(\nu)$ and $t(\nu)$ is a t distribution with degrees of freedom ν.

2.1 Equivalence Test of Linear Trend

The primary focus of this article is the test of equivalence, the null and alternative hypotheses are expressed as

$$\mathrm{H}_0 : \beta_1 \leq \Delta_L \text{ or } \Delta_U \leq \beta_1 \text{versus } \mathrm{H}_1 : \Delta_L < \beta_1 < \Delta_U, \tag{5}$$

where Δ_L and Δ_U are a priori constants that represent the minimal range for declaring equivalence effect size. The hypotheses with asymmetric equivalence thresholds can be readily rewritten in terms of symmetric equivalence bounds as

$$\mathrm{H}_0 : \beta_1^* \leq -\Delta \text{ or } \Delta \leq \beta_1^* \text{ versus } \mathrm{H}_1 : -\Delta < \beta_1^* < \Delta, \tag{6}$$

where $\beta_1^* = \beta_1 - \Delta_M$, $\Delta_M = (\Delta_L + \Delta_U)/2$, and $\Delta = (\Delta_U - \Delta_L)/2$. An important scenario is to detect a negligible trend by setting $\Delta_U = \Delta$ and $\Delta_L = -\Delta$ so that $\Delta_M = 0$ for a bound Δ.

For the given value of the predictor quantity SSX, it is essential to note that

$$T_S = \frac{\hat{\beta}_1 - \Delta_M}{(\hat{\sigma}^2/SSX)^{1/2}} \sim t(\nu, \lambda_S), \tag{7}$$

where $t(\nu, \lambda_S)$ is the noncentral t distribution with degrees of freedom ν and noncentrality parameter $\lambda_S = (\beta_1 - \Delta_M)/(\sigma^2/SSX)^{1/2}$. To claim the slope coefficient β_1 is within the interval (Δ_L, Δ_U), a natural rejection region to the null hypothesis is

$$\{\tau_{SL} < T_S < \tau_{SU}\},$$

where the two critical values τ_{SL} and τ_{SU} are chosen to simultaneously attain the nominal Type I error rate

$$P\{\tau_{SL} < T_S < \tau_{SU} | \beta_1 = \Delta_L\} = \alpha \text{ and } P\{\tau_{SL} < T_S < \tau_{SU} | \beta_1 = \Delta_U\} = \alpha.$$

Following the properties of a noncentral t distribution as in Johnson, Kotz and Balakrishnan [20], it can be shown that the two conditions can be simultaneously satisfied by the choice of critical values $\tau_{SL} = -\tau_S$ and $\tau_{SU} = \tau_S$ where $\tau_S > 0$. Hence, the rejection region is of the form

$$AH_S = \{-\tau_S < T_S < \tau_S\}, \tag{8}$$

where τ_S is determined by the condition $P\{-\tau_S < T_S < \tau_S | \beta_1 = \Delta_L\} = \alpha$ or $P\{-\tau_S < T_S < \tau_S | \beta_1 = \Delta_U\} = \alpha$. Note that the error variance is generally unknown and the exact distribution of T_S cannot be specified. Following the suggestion in Anderson and Hauck [15], a feasible and accurate approach is to find the critical value τ_S through the approximation $T_S \dot{\sim} T + \hat{\lambda}_S$ where $T \sim t(\nu)$, $\hat{\lambda}_S = \Delta/(\hat{\sigma}^2/SSX)^{1/2}$, and

$$P\{-\tau_S < T + \hat{\lambda} < \tau_S\} = \alpha. \tag{9}$$

Thus, the optimal quantity τ_S can be computed by a simple iterative search. Note that the critical value τ_S is a function of $\alpha, \Delta, N, \hat{\sigma}^2$, and SSX. It does not have an explicit analytic expression and requires a computer program to calculate the actual value. An efficient algorithm is developed for computing the critical value and rejection region for the suggested procedure. Also, the p-value associated with the observed slope estimate $\hat{\beta}_{1O}$ can be calculated as

$$p\text{-value} = P\{-|T_O| - \hat{\lambda}_S < T < |T_O| - \hat{\lambda}_S\}, \tag{10}$$

where $T_O = (\hat{\beta}_{1O} - \Delta_M)/(\hat{\sigma}^2/SSX)^{1/2}$. It is apparent that the p-value is computationally easier to obtain than the critical value.

Note that similar discussion was described in Anderson and Hauck [15] for testing two-group mean equivalence. Because of the computational ease of the p-value, they recommend the p-value approach to conclude the decision. Hence, they did not address the calculation and implementation issues of the rejection region and corresponding power function. Accordingly, the sample size procedure for mean equivalence in Hauck and Anderson [16] is less transparent and cannot be readily adopted as a general tool in linear regressions. Moreover, the Anderson and Hauck procedure has an unbounded rejection region as other more powerful tests. The counterintuitive rejection of nonequivalence with arbitrarily large values of sample variance has been debated extensively in Berger and Hsu [10] and the discussions therein. As a constructive response, they proposed to specify an upper bound on the sample variance beyond which the null hypothesis will never be rejected. Moreover, unlike the TOST, the advantage of the Anderson and Hauck procedure in the Type I error protection for small sample sizes and tight equivalence bounds should also be taken into consideration. The contrasting behavior of the two test procedures is also demonstrated in the subsequent numerical examples.

2.2 Equivalence Test of Mean Response

The equivalence appraisal can also be applied to the mean response $\mu = \beta_0 + X\beta_1$ at a focal predictor value X_F. The null and alternative hypotheses are presented as

$$\mathrm{H}_0 : \mu \le \Delta_L \text{ or } \Delta_U \le \mu \text{ versus } \mathrm{H}_1 : \Delta_L < \mu < \Delta_U, \tag{11}$$

where Δ_L and Δ_U are a priori constants that represent the threshold range for declaring practical equivalence. With the least squares estimators $(\hat{\beta}_0, \hat{\beta}_1)$ of (β_0, β_1), the linear estimator $\hat{\mu} = \hat{\beta}_0 + X_F\hat{\beta}_1$ has the distribution

$$\hat{\mu} \sim N(\mu, \sigma^2 H_M), \tag{12}$$

where $H_M = 1/N + (X_F - \overline{X})^2/SSX$. It is useful to note that

$$T_M = \frac{\hat{\mu} - \Delta_M}{(\hat{\sigma}^2 H_M)^{1/2}} \sim t(\nu, \lambda_M), \tag{13}$$

where the noncentrality parameter $\lambda_M = (\mu - \Delta_M)/(\sigma^2 H_M)^{1/2}$ and $\Delta_M = (\Delta_L + \Delta_U)/2$.

Following the same principle for slope coefficient assessment, a potential rejection region to the null hypothesis is of the form

$$AH_M = \{-\tau_M < T_M < \tau_M\}, \tag{14}$$

where the critical value τ_M is chosen to attain the nominal Type I error rate when $\mu = \Delta_L$ and Δ_U. The proposed approach is to find the critical value through the approximate evaluation

$$P\{-\tau_M < T + \hat{\lambda}_M < \tau_M\} = \alpha \tag{15}$$

where $T \sim t(\nu)$, $\hat{\lambda}_M = \Delta/(\hat{\sigma}^2 H_M)^{1/2}$, and $\Delta = (\Delta_U - \Delta_L)/2$. Note that the critical value τ_M is a function of α, Δ, N, $\hat{\sigma}^2$, and H_M. Moreover, an iterative algorithm is required to compute the critical value.

2.3 A Numerical Example

The numerical details for the equivalence tests of slope coefficient and mean response are demonstrated with the data of training study described in Table 6.1 of Huitema [21] about the relation between the response variable (Y: achievement) and the predictor variable (X: aptitude) for three types of training program.

For the first training group with $N = 10$, the sample means of the predictor and response variables are $\overline{X} = 52.00$ and $\overline{Y} = 30.00$, respectively. Moreover, the least squares estimates of the linear regression line between achievement and aptitude measurements are obtained as $\{\hat{\beta}_0,\ \hat{\beta}_1\} = \{4.1033,\ 0.4980\}$, and the sample variance of error is $\hat{\sigma}^2 = 70.5615$. For illustration, an equivalence test of slope coefficient is performed in terms of H_0: $\beta_1 \le 0.25$ or $0.75 \le \beta_1$ versus H_1: $0.25 < \beta_1 < 0.75$ ($\Delta_M = 0.50$ and $\Delta = 0.25$). With $SSX = 2014.00$ and $\alpha = 0.05$, the test statistic and critical value are computed as $T_S = -0.0106$ and $\tau_S = 0.1598$, respectively. Thus, the nonequivalence null hypothesis is rejected at the significance level 0.05. The conclusion indicates that the slope coefficient is essentially equivalent to 0.50 with no more than 0.25 difference.

The equivalence test of mean response can also be performed with the estimated mean response $\hat{\mu} = 29.0040$ at $X_F = 50$. Using $\Delta_M = 29$ and $\Delta = 4$, the equivalence test of mean response is conducted in terms of H_0: $\mu \le 25$ □or $33 \le \mu$ versus H_1: $25 < \mu < 33$. The test statistic and critical value can be computed as $T_M = 0.0015$ and $\tau_M = 0.1966$, respectively for $\alpha = 0.05$. Hence, the nonequivalence null hypothesis is rejected at the significance level 0.05. The analysis suggests that the mean response

at $X_F = 50$ is nearly within a bound of 4 around 29. Moreover, it can be shown that the resulting rejection regions of the TOST procedures are empty sets and there is no chance to reject the nonequivalence null hypothesis of the slope coefficient and mean response. Apparently, the TOST approach may not be a reliable procedure when the sample size is small, especially for a tight equivalence range. Such deficiency agrees with the explication of TOST for assessing mean equivalence in Schuirmann [12].

2.4 Power and Sample Size Calculations

When planning and conducting a research, the actual values of the continuous measurements of response and predictor variable for each subject are available only after the observations are obtained. In addition to the randomness of normal responses, the stochastic nature of predictor variables has to be taken into account in power analysis under the random and unconditional context in linear regression study. A useful and convenient framework is to assume the continuous predictor variables $\{X_i, i = 1, \ldots, N\}$ have the independent and identical normal distribution $N(\mu_X, \sigma_X^2)$ as in Shieh [22, 23] within the context of ANCOVA.

Under the prescribed stochastic consideration of $\{X_i, i = 1, \ldots, N\}$, it can be readily established that $K = SSX/\sigma_X^2 \sim \chi^2(\kappa)$ where $\kappa = N - 1$. The power function of the equivalence procedure for slope coefficient can be expressed as

$$\Pi_S = P\{-\tau_S < T_S < \tau_S | \Delta_L < \beta_1 < \Delta_U\}. \tag{16}$$

Note that the critical value τ_S depends on the two quantities $\hat{\sigma}^2$ and SSX. With $\hat{\sigma}^2 = \sigma^2(V/\nu)$ and $Hs = 1/SSX = 1/(\sigma_X^2 K)$, the power function Π_S can be rewritten as

$$\Pi_S = E_{(K,V)}[\Phi(B_S) - \Phi(A_S)], \tag{17}$$

where $B_S = (\Delta_M - \beta_1)/(\sigma^2 Hs)^{1/2} + \tau_S(V/\nu)^{1/2}$, $A_S = (\Delta_M - \beta_1)/(\sigma^2 Hs)^{1/2} - \tau_S(V/\nu)^{1/2}$, $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, and the expectation $E_{(K,V)}$ is taken with respect to the chi-square distributions of K and V.

Under the random predictor framework, the normality assumption implies that

$$T_X = \frac{\bar{X} - X_F}{(\widehat{\sigma}_X^2/N)^{1/2}} \sim t(\kappa, \lambda_X), \tag{18}$$

where $\hat{\sigma}_X^2 = SSX/\kappa$ and $\lambda_X = (\mu_X - X_F)/(\sigma_X^2/N)^{1/2}$ Also, the power function of the equivalence procedure for mean response is of the form

$$\Pi_M = P\{-\tau_M < T_M < \tau_M | \Delta_L < \mu < \Delta_U\}. \tag{19}$$

In this case, the critical value τ_M depends on the two terms $\hat{\sigma}^2$ and H_M. With $\hat{\sigma}^2 = \sigma^2(V/\nu)$ and $H_M = 1/N + T_X^2/(\kappa N)$, it follows that the power function Π_M

can be expressed as

$$\Pi_M = E_{(TX,V)}[\Phi(B_M) - \Phi(A_M)], \tag{20}$$

where $B_M = (\Delta_M - \mu)/(\sigma^2 H_M)^{1/2} + \tau_M(V/\nu)^{1/2}$, $A_M = (\Delta_M - \mu)/(\sigma^2 H_M)^{1/2} - \tau_M(V/\nu)^{1/2}$, and $E_{(TX,V)}$ is taken with respect to the joint distribution of T_X and V.

The prescribed power functions Π_S and Π_M for slope coefficient and mean response involve a mixture of noncentral t distributions through the distribution K and T_X of the predictor variables, respectively. It is appealing to simplify these power functions because of computational complexity. Under the normal assumption $N(\mu_X, \sigma_X^2)$ for the predictors $\{X_i, i = 1, \ldots, N\}$, the standard results show that $E[\overline{X}] = \mu_X$ and $E[SSX] = \kappa\sigma_X^2$. Hence, an approximation of unconditional distribution can be obtained for the test statistic $T_S \dot{\sim} t(\nu, \lambda_{SA})$ where $\lambda_{SA} = (\beta_1 - \Delta_M)/(\sigma^2 H_{SA})^{1/2}$ and $H_{SA} = 1/(\kappa\sigma_X^2)$. It yields a simplified power function for the equivalence test of linear trend

$$\Pi_{SA} = P\{-\tau_S < t(\nu, \lambda_{SA}) < \tau_S\}. \tag{21}$$

Moreover, following similar arguments, the test statistic of mean response has the approximate distribution $T_M \dot{\sim} t(\nu, \lambda_{MA})$ where $\lambda_{MA} = (\mu - \Delta_M)/(\sigma^2 H_{MA})^{1/2}$ and $H_{MA} = 1/N + (\mu_X - X_F)^2/(\kappa\sigma_X^2)$. Then, an approximate power function for the equivalence test of mean response is denoted by

$$\Pi_{MA} = P\{-\tau_M < t(\nu, \lambda_{MA}) < \tau_M\}. \tag{22}$$

The approximate power functions of the equivalence procedures provide computational shortcuts to the exact formulas. The simple formulations can be readily implemented with the embedded probability functions of a noncentral t distribution in standard software systems. On the other hand, the prescribed analytic justifications provide statistical support for the exact power functions. An immediate application of the power functions is to compute optimal sample sizes needed for the equivalence procedure to attain the specified power under the designated model configurations. The fundamental discrepancy between the exact and simplified power and sample size calculations will be further assessed in the succeeding numerical investigations.

2.5 Numerical Assessments

As an exemplifying framework, the model configurations follow that of the prescribed training study in Huitema [21]. Accordingly, the sample estimated of regression coefficients and variance component of the first training group are designated the working configurations: $\{\beta_0, \beta_1\} = \{4.1033, 0.4980\}$, and $\sigma^2 = 70.5615$, respectively. The mean and variance of the normal predictors are chosen as $\{\mu_X, \sigma_X^2\} = \{52.00, 223.7778\}$. The equivalence thresholds (Δ_L, Δ_U) are defined as $\Delta_L = \Delta_M - \Delta$, $\Delta_U = \Delta_M + \Delta$, and various magnitudes of Δ_M and Δ are evaluated. For the equivalence tests of linear trend, the selected values are $\Delta_M = 0.5$ with $\Delta = 0.2, 0.3$, and

0.4. The equivalence tests of mean response are examined at $X_F = 50$ with $\Delta_M = 29$ for $\mu = 29.0040$ under three equivalence bounds $\Delta = 4$, 5, and 6.

With these specifications, the required sample sizes of both exact and approximate methods were computed for the chosen power value $1 - \beta = 0.80$ and significance level $\alpha = 0.05$. The estimated sample sizes for the equivalence tests of linear trend and mean response are presented in Table 1. Note that the resulting sample sizes cover a reasonable range of magnitudes without being unrealistic or excessively large. More importantly, the estimated sample sizes of the exact approach are consistently larger than or equal to those of the approximate procedure for all 6 cases. For ease comparing the accuracy of power functions, the estimated power or attained power are also summarized in Table 1. Because of the underlying metric of integer sample sizes, the estimated values of both exact and approximate procedures are marginally larger than the nominal level for all cases.

In the second stage, Monte Carlo simulation studies were performed to justify the performance of power and sample size calculations. With the computed sample sizes, parameter configurations, and nominal alpha level, estimates of the true power were computed via Monte Carlo simulation of 10,000 independent data sets. For each replicate, the sample size N predictor values were generated from the selected normal distributions. The outcome values of predictor variables are then designated to determine the mean responses for generating the normal responses with the specified linear regression model. Next, the equivalence test statistics were computed and the simulated power was the proportion of the 10,000 replicates whose null hypothesis was rejected at the significance level 0.05. Accordingly, the adequacy of the approximate and exact sample size procedures is determined by the error (= simulated power – estimated power) between the simulated power of Monte Carlo study and the estimated power computed from analytic power function. The simulated power and error are also presented in Table 1.

The results reveal that the exact approaches are extremely accurate because the associated errors of the 6 cases are all within the small range of –0.0055 to 0.0075. Accordingly, there exists a close agreement between the simulated power and the estimated power of the exact approaches for these settings. On the other hand, the simulated powers for the approximate methods are constantly less than the estimated powers. Specifically, the resulting errors are {–0.0167, –0.0210, –0.0306} and {–0.0057, –0.0069, –0.0177} for the linear trend and mean response, respectively. Although some of the differences are not substantial, it implies that the approximate power functions do not give reliable results for small sample sizes. In short, the adequacy of the approximate power and sample size calculations varies with model configurations. It is clear that the exact techniques are more reliable and accurate than the approximate methods for all cases of linear trend and mean response considered here.

3 Two Regression Lines

The two-group nonparallel simple linear regression model is expressed as

$$Y_{1i} = \beta_{01} + X_{1i}\beta_{11} + \varepsilon_{1i} \text{ and } Y_{2j} = \beta_{02} + X_{2j}\beta_{12} + \varepsilon_{2j}, \tag{23}$$

Table 1 Computed sample size, estimated power, and simulated power for the Anderson and Hauck test of linear trend and mean response at $X_F = 50$ when Type I error $\alpha = 0.05$ and nominal power $1 - \beta = 0.80$

Parameter	Δ_M	Δ	(Δ_L, Δ_U)	Exact approach				Approximate method			
				N	Simulated power	Estimated power	Error	N	Simulated power	Estimated power	Error
$\beta_1 = 0.4980$	0.5	0.2	(0.3, 0.7)	73	0.8141	0.8066	0.0075	70	0.7852	0.8019	− 0.0167
		0.3	(0.2, 0.8)	35	0.8039	0.8094	− 0.0055	33	0.7851	0.8061	− 0.0210
		0.4	(0.1, 0.9)	22	0.8269	0.8223	0.0046	20	0.7830	0.8136	− 0.0306
$\mu = 29.0040$	29	4	(25, 33)	41	0.8029	0.8020	0.0009	41	0.8071	0.8128	− 0.0057
		5	(24, 34)	27	0.8023	0.8001	0.0022	27	0.8086	0.8155	− 0.0069
		6	(23, 35)	20	0.8123	0.8144	− 0.0021	19	0.7893	0.8070	− 0.0177

where ε_{1i} and ε_{2j} are *iid* $N(0, \sigma^2)$ random variables, $i = 1, \ldots, N_1$, and $j = 1, \ldots, N_2$. Note that a traditional ANCOVA model assumes that the regression slopes are equivalent $\beta_{11} = \beta_{12}$. Accordingly, a test of slope equality is generally required to justify the use of ANCOVA.

Standard results that the least squares estimators $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ of slope coefficients β_{11} and β_{12} have the following distributions

$$\hat{\beta}_{11} \sim N(\beta_{11}, \sigma^2/SSX_1) \text{ and } \hat{\beta}_{12} \sim N(\beta_{12}, \sigma^2/SSX_2),$$

where $SSX_1 = \sum_{i=1}^{N_1} (X_{1i} - \overline{X})^2, SSX_2 = \sum_{j=1}^{N_2} (X_{2j} - \overline{X}_2)^2$, $\overline{X}_1 = \sum_{i=1}^{N_1}, X_{1i}/N_1$ and $\overline{X}_2 = \sum_{i=1}^{N_2} X_{2i}/N_2$. The difference of two slope estimators has the distribution

$$\hat{\beta}_D = \hat{\beta}_{11} - \hat{\beta}_{12} \sim N\{\beta_D, \sigma^2 H_{DS}\}, \tag{24}$$

where $\beta_D = \beta_{11} - \beta_{12}$ and $H_{DS} = 1/SSX_1 + 1/SSX_2$. In this case, $\hat{\sigma}^2 = SSE/\nu_D$ is the usual unbiased estimator of σ^2 and $V = SSE/\sigma^2 \sim \chi^2(\nu_D)$ where SSE is the error sum of squares and $\nu_D = N_1 + N_2 - 4$.

To detect the difference between two slope coefficients in terms of H_0: $\beta_D = \beta_{D0}$ versus H_1: $\beta_D \neq \beta_{D0}$, the test statistic has the form

$$T_{DS0} = \frac{\hat{\beta}_D - \beta_{D0}}{(\hat{\sigma}^2 H_{DS})^{1/2}} \tag{25}$$

The null hypothesis is rejected at the significance level α if

$$|T_{DS0}| > t_{\nu_D, \alpha/2} \tag{26}$$

3.1 Equivalence Test of Trend Effect

To conduct equivalence test of trend effect or slope difference, the null and alternative hypotheses are expressed as

$$H_0 : \beta_D \leq \Delta_L \text{ or } \Delta_U \leq \beta_D \text{ versus } H_1 : \Delta_L < \beta_D < \Delta_U, \tag{27}$$

where Δ_L and Δ_U are a priori constants that denote the minimal magnitude for declaring equivalence for trend effect. Under the model assumption, it follows that

$$T_{DS} = \frac{\hat{\beta}_D - \Delta_M}{(\hat{\sigma}^2 H_{DS})^{1/2}} \sim t(\nu_D, \lambda_{DS}), \tag{28}$$

where the noncentrality parameter $\lambda_{DS} = (\beta_D - \Delta_M)/(\sigma^2 H_{DS})^{1/2}$ and $\Delta_M = (\Delta_L + \Delta_U)/2$. To justify the slope difference β_D is within the interval (Δ_L, Δ_U), a feasible

rejection region to the null hypothesis is

$$AH_{DS} = \{-\tau_{DS} < T_{DS} < \tau_{DS}\}, \tag{29}$$

where the critical value τ_{DS} is chosen to simultaneously attain the nominal Type I error rate when $\beta_D = \Delta_L$ and Δ_U. In practice, the exact distribution of T_{DS} is practically unknown and the critical value τ_{DS} can be determined through the approximation

$$P\{-\tau_{DS} < T + \hat{\lambda}_{DS} < \tau_{DS}\} = \alpha, \tag{30}$$

where $T \sim t(\nu_D)$, $\hat{\lambda}_{DS} = \Delta/(\hat{\sigma}^2 H_{DS})^{1/2}$, and $\Delta = (\Delta_U - \Delta_L)/2$. The optimal quantity τ_{DS} is a function of α, Δ, N_1, N_2, $\hat{\sigma}^2$, and H_{DS}. Although the critical value does not have a closed-form expression, it can be computed by a simple iterative search.

As emphasized in Huitema [21], Kutner et al. [24], Rencher and Schaalje [25], and related texts of research methods, the traditional ANCOVA assumes that the slope coefficients associating the predictor variables with the response variables are the same for each treatment group. The assertion of homogeneous regression slopes implies a lack of interaction effects between a categorical moderator and a continuous predictor in moderation study. Note that the conventional difference test purports to show the regression lines are nonparallel. Hence, the suggested equivalence procedure for trend effect is more appropriate for supporting the equality or comparability of slope coefficients assumption in ANCOVA.

3.2 Equivalence Test of Simple Effect

A related and practical scheme for comparing two regression lines is to assess the difference between two mean responses at a designated predictor value. The simple effect or the mean response difference between two regression lines at X_F is defined as

$$\mu_D = \mu_1 - \mu_2 = (\beta_{01} - \beta_{02}) + X_F(\beta_{11} - \beta_{12}) \tag{31}$$

The equivalence test of simple effect is conducted under the null and alternative hypotheses:

$$\mathrm{H}_0 : \mu_D \leq \Delta_L \text{ or } \Delta_U \leq \mu_D \text{ versus } \mathrm{H}_1 : \Delta_L < \mu_D < \Delta_U, \tag{32}$$

where Δ_L and Δ_U are a priori constants that represent the minimal threshold for declaring essential equivalence.

Using the least squares estimators $\{\hat{\beta}_{01}, \hat{\beta}_{11}\, \hat{\beta}_{02}, \hat{\beta}_{12}\}$ of for the intercept and slope coefficients $\{\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}\}$, the estimated mean response $\hat{\mu}_1$ and $\hat{\mu}_2$ for mean values $\mu_1 = \beta_{01} + X\beta_{11}$ and $\mu_2 = \beta_{02} + X\beta_{12}$ at a specified value X_F are

$$\hat{\mu}_1 = \hat{\beta}_{01} + X_F\hat{\beta}_{11} \text{ and } \hat{\mu}_2 = \hat{\beta}_{02} + X_F\hat{\beta}_{12}$$

respectively. A natural and unbiased estimator of μ_D is $\hat{\mu}_D = \hat{\mu}_1 - \hat{\mu}_2$ and

$$\hat{\mu}_D \sim N(\mu_D, \sigma^2 H_{DM}), \tag{33}$$

where $H_{DM} = 1/N_1 + 1/N_2 + (X_F - \overline{X}_1)^2/SSX_1 + (X_F - \overline{X}_2)^2/SSX_2$. It is important to note under the model assumption that

$$T_{DM} = \frac{\hat{\mu}_D - \Delta_M}{(\hat{\sigma}^2 H_{DM})^{1/2}} \sim t(\nu_D, \lambda_{DM}), \tag{34}$$

where the noncentrality parameter $\lambda_{DM} = (\mu_D - \Delta_M)/(\sigma^2 H_{DM})^{1/2}$ and $\Delta_M = (\Delta_L + \Delta_U)/2$. To evaluate whether the simple effect μ_D is within the interval (Δ_L, Δ_U), the suggested rejection region is

$$AH_{DM} = \{-\tau_{DM} < T_{DM} < \tau_{DM}\}, \tag{35}$$

where the critical value τ_{DM} is chosen to simultaneously attain the nominal Type I error rate when $\mu_D = \Delta_L$ and Δ_U. The assessments can be calculated through the approximation

$$P\{-\tau_{DM} < T + \hat{\lambda}_{DM} < \tau_{DM}\} = \alpha, \tag{36}$$

where $T \sim t(\nu_D)$, $\hat{\lambda}_{DM} = \Delta/(\hat{\sigma}^2 H_{DM})^{1/2}$, and $\Delta = (\Delta_U - \Delta_L)/2$. The optimal quantity τ_{DM} is a function of α, Δ, N_1, N_2, $\hat{\sigma}^2$, and H_{DM}, and it needs to be calculated by an iterative search algorithm.

It should be noted that the equivalence analysis of simple effect or response difference between two linear regression lines is closely related to the Johnson–Neyman problem of Johnson and Neyman [26] and Potthoff [27]. The Johnson–Neyman regions of significance and non-significance are identified with the conclusion to reject or the failure to reject the conventional hypothesis of no difference between mean responses. Technical illustrations and implications can be found in Hunka [28], Rogosa [29], and Spiller, et al. [30], among others. Contrastly, the proposed equivalence test of simple effect can be used to identify the regions of equivalence and nonequivalence or the ranges of predictor values that the simple effect is equivalent and nonequivalent.

3.3 An Application

The prescribed example about training study in Table 6.1 of Huitema [21] is utilized to demonstrate the suggested equivalence testing of trend and simple effects between the first two treatments. In addition to the summary information of the first group, the second group of training type has $N_2 = 10$, $\overline{Y}_2 = 39.0000$ and $\overline{X}_2 = 47.0000$, and $SSX_2 = 1798.00$. The regression coefficient estimates are $\{\hat{\beta}_{02}, \hat{\beta}_{12}\} = \{15.1863,\ 0.5067\}$ and the sample variance of error is $\hat{\sigma}_2^2 = 54.3025$. It is readily obtained that $\hat{\beta}_D = \hat{\beta}_{11} - \hat{\beta}_{12} = -0.0087$ and the pooled sample variance is $\hat{\sigma}^2 = 62.4320$. The

equivalence hypothesis testing of trend effect is presented as H_0: $\beta_D \leq -0.25$ or $0.25 \leq \beta_D$ versus H_1: $-0.25 < \beta_D < 0.25$ ($\Delta_M = 0$ and $\Delta = 0.25$). For $\nu = 16$ and $\alpha = 0.05$, the test statistic $T_{DS} = -0.0338$ and the critical value $\tau_{DS} = 0.1048$. Hence, the nonequivalence null hypothesis is rejected at the significance level 0.05. It suggests that the slope coefficient is virtually equivalent and their difference is within the range (–0.25, 0.25).

It is of practical importance to assess the simple effect or the mean response difference between two regression lines. At the particular predictor value $X_F = 50$, the mean response difference is computed as $\hat{\mu}_D = \hat{\mu}_1 - \hat{\mu}_2 = -11.5161$. For illustration, the equivalence thresholds is set as $\Delta_M = -11$ and $\Delta = 5$ and the equivalence test of simple effect is conducted for the hypotheses H_0: $\mu_D \leq -16$ or $-6 \leq \mu_D$ versus H_1: $-16 < \mu_D < -6$. With $\nu = 16$ and $\alpha = 0.05$, the test statistic and critical value can be obtained as $T_{DM} = -0.1436$ and $\tau_M = 0.1684$, respectively. Consequently, the nonequivalence null hypothesis is rejected at the significance level 0.05 and the mean response difference is practically –11 with the threshold of 5 at $X_F = 50$. In view of the limited features of available software packages, computer programs are developed to facilitate the usage of the proposed equivalence procedures for trend and simple effects.

3.4 Power and Sample Size Calculations

In order to elucidate the critical notion of accommodating the distributional properties of the predictor variables, the continuous covariate variables $\{X_{1i}, i = 1, \ldots, N_1\}$ and $\{X_{2j}, j = 1, \ldots, N_2\}$ are assumed to have the independent normal distributions $N(\mu_{X1}, \sigma^2_{X1})$ and $N(\mu_{X2}, \sigma^2_{X2})$, respectively. It can be readily established that $K_1 = SSX_1/\sigma^2_{X1} \sim \chi^2(\kappa_1)$ and $K_2 = SSX_2/\sigma^2_{X2} \sim \chi^2(\kappa_2)$ where $\kappa_1 = N_1 - 1$ and $\kappa_2 = N_2 - 1$.

Under the unconditional setting, the power function for trend effect is expressed as

$$\Pi_{DS} = P\{-\tau_{DS} < T_{DS} < \tau_{DS} | \Delta_L < \beta_D < \Delta_U\}. \tag{37}$$

Note that the critical value τ_{DS} depends on the two statistics $\hat{\sigma}^2$ and H_{DS}. With $\hat{\sigma}^2 = \sigma^2(V/\nu)$ and $H_{DS} = 1/(\sigma^2_{X1} K_1) + 1/(\sigma^2_{X2} K_2)$, the power function Π_{DS} can be rewritten as

$$\Pi_{DS} = E_{(K1, K2, V)}[\Phi(B_{DS}) - \Phi(A_{DS})], \tag{38}$$

where $B_{DS} = (\Delta_M - \beta_D)/(\sigma^2 H_{DS})^{1/2} + \tau_{DS}(V/\nu_D)^{1/2}$, $A_{DS} = (\Delta_M - \beta_D)/(\sigma^2 H_{DS})^{1/2} - \tau_{DS}(V/\nu_D)^{1/2}$, and $E_{(K1, K2, V)}$ is taken with respect to the joint distribution of K_1, K_2 and V.

Moreover, the normality assumptions of predictor variables imply that

$$T_{Xg} = \frac{\overline{X}_g - X_F}{(\hat{\sigma}^2_{X_g}/N_g)^{1/2}} \sim t(\kappa_g, \lambda_{X_g}) \tag{39}$$

where $\hat{\sigma}^2_{Xg} = SSX_g/\kappa_g$ and $\lambda_{Xg} = (\mu_{Xg} - X_F)/(\sigma^2_{Xg}/N_g)^{1/2}$ for $g = 1$ and 2. Following the prescribed power function Π_{DS}, the power function for mean response difference is presented as

$$\Pi_{DM} = P\{-\tau_{DM} < T_{DM} < \tau_{DM} | \Delta_L < \mu_D < \Delta_U\}. \tag{40}$$

Note that the critical value τ_{DM} depends on the two terms $\hat{\sigma}^2$ and H_{DM}. With $\hat{\sigma}^2 = \sigma^2(V/\nu_D)$, $H_{DM} = 1/N_1 + 1/N_2 + T^2_{X1}/(\kappa_1 N_1) + T^2_{X2}/(\kappa_2 N_2)$, the power function has the alternative form

$$\Pi_{DM} = E_{(TX1, TX2, V)}[\Phi(B_{DM}) - \Phi(A_{DM})], \tag{41}$$

where $B_{DM} = (\Delta_M - \mu_D)/(\sigma^2 H_{DM})^{1/2} + \tau_{DM}(V/\nu_D)^{1/2}$, $A_{DM} = (\Delta_M - \mu_D)/(\sigma^2 H_{DM})^{1/2} - \tau_{DM}(V/\nu_D)^{1/2}$, and $E_{(TX1, TX2, V)}$ is taken with respect to the joint distribution of T_{X1}, T_{X2} and V.

It is also temping to simplify the unconditional distributions for the equivalence test statistics for comparing slope coefficients and mean responses. Conceivably, a straightforward approach is to replace the two means $\{\overline{X}_1, \overline{X}_2\}$ and sum of squares $\{SSX_1, SSX_2\}$ with the corresponding expected values $E[\overline{X}_1] = \mu_{X1}$, $E[\overline{X}_2] = \mu_{X2}$, $E[SSX_1] = \kappa_1\sigma^2_{X1}$, and $E[SSX_2] = \kappa_2\sigma^2_{X2}$. Thus, an approximate power function for the equivalence test of trend effect is

$$\Pi_{DSA} = P\{-\tau_{DS} < t(\nu, \lambda_{DSA}) < \tau_{DS}\}, \tag{42}$$

where $\lambda_{DSA} = (\beta_D - \Delta_M)/(\sigma^2 H_{DSA})^{1/2}$ and $H_{DSA} = 1/(\kappa_1\sigma^2_{X1}) + 1/(\kappa_2\sigma^2_{X2})$. Moreover, the power function of equivalence test of simple effect is expressed as

$$\Pi_{DMA} = P\{-\tau_{DM} < t(\nu, \lambda_{DMA}) < \tau_{DM}\}, \tag{43}$$

where $\lambda_{DMA} = (\mu_D - \Delta_M)/(\sigma^2 H_{DMA})^{1/2}$ and $H_{DMA} = 1/N_1 + 1/N_2 + (\mu_{X1} - X_F)^2/(\kappa_1\sigma^2_{X1}) + (\mu_{X2} - X_F)^2/(\kappa_2\sigma^2_{X2})$. Empirical examinations will be conducted to demonstrate the critical differences between the exact and approximate power functions using different levels of information of predictor variables.

3.5 Numerical Investigations

The model configurations of the first two groups of the training study in Huitema [21] provide a convenient framework for the subsequent simulation study of trend effect and simple effect. For illustration, the key statistics of response and predictor variables are treated as population parameters as potential settings of future investigations for power calculations and sample size determinations. Specifically, the regression coefficients are $\{\beta_{01}, \beta_{11}\} = \{4.1033, 0.4980\}$, $\{\beta_{02}, \beta_{12}\} = \{15.1863, 0.5067\}$, and common error variance $\sigma^2 = 62.4320$. The means and variances of the two predictor variables are $\{\mu_{X1}, \sigma^2_{X1}\} = \{52.00, 223.7778\}$ and $\{\mu_{X2}, \sigma^2_{X2}\} = \{47.00, 199.7778\}$.

Similar to the prescribed scenario of linear trend and mean response, numerical investigations contain the determination of optimal sample sizes and the simulation study of power calculations. Through the empirical examinations, the Type I error rate and nominal power are fixed as $\alpha = 0.05$ and $1 - \beta = 0.80$, respectively. First, the trend effect or the slope difference between two regression lines is $\beta_D = -0.0087$. Thus, the equivalence tests of trend effect have $\Delta_M = 0$ and $\Delta = 0.2, 0.3$, and 0.4 for the equivalence bounds. Second, the mean response of the two levels of treatment at $X_F = 50$ are $\mu_1 = 29.0040$ and $\mu_2 = 40.5200$, respectively, and their difference is $\mu_D = -11.5161$. Accordingly, the equivalence tests of simple effect are performed for $\Delta_M = -11$ and $\Delta = 4, 5$, and 6. The optimal sample sizes of both exact approach and approximate method were determined for the chosen power value and significance level with balanced and unbalanced structures $r = N_1/N_2 = 1$ and 2. The computed sample sizes for the equivalence tests of trend effect and simple effect are presented in Tables 2 and 3, respectively. The results suggest the general pattern that the approximate formulas tend to give smaller sample sizes than the exact techniques. Balanced designs require fewer samples to achieve the nominal power than the unbalanced structures. Also, the computed sample size decreases with increasing threshold bound Δ.

To elucidate the accuracy of sample size calculations, Monte Carlo simulation study of 10,000 replications were conducted to obtain the simulated powers and they are compared to the estimated powers for the optimal sample sizes. These power values and associated errors are also presented in the tables. As can been from the reported deviations, the exact approaches of trend effect and simple effect maintain small errors in power computations. Whereas the approximate methods are not as good as the exact counterparts and their performance deteriorates as the sample size decreases. Specifically, the two errors associated with $\Delta = 0.4$ are $\{-0.0301, -0.0360\}$ and $\{-0.0172, -0.0157\}$ in Tables 2 and 3, respectively. The overall usefulness of the approximate methods is affected by the undesirable properties of underestimation of sample sizes and over-calculation of power levels. According to the findings, the exact power functions and sample size procedures are recommended for general use. The implementation of the suggested power evaluation and sample size determination involves specialized programs not currently available in prevailing statistical packages. Thus, the accompanying computer algorithms are presented for conducting the suggested power and sample size calculations.

4 Conclusions

The concept and theory of equivalence have been widely practiced in pharmaceutical sciences and related medical fields. Equivalence testing procedures are also potentially useful in behavioral and psychological sciences. The technical intuition and computational simplicity of TOST provide an important motivation to apply appropriate statistical tools for equivalence assessment, rather than the traditional hypothesis tests that purport to detect whether treatment groups significantly differ from one another. Despite the ready applicability, the TOST is generally conservative and the true Type I error rate can be substantially less than the nominal level for close equivalence bounds and small sample sizes. In contrast, the Anderson and Hauck procedure and other more

Table 2 Computed sample size, estimated power, and simulated power for the Anderson and Hauck test of trend effect when Type I error $\alpha = 0.05$ and nominal power $1 - \beta = 0.80$

Parameter	Δ_M	Δ	(Δ_L, Δ_U)	Exact approach				Approximate method			
				(N_1, N_2)	Simulated power	Estimated power	Error	(N_1, N_2)	Simulated power	Estimated power	Error
$\beta_D = -0.0087$	0	0.2	(− 0.2, 0.2)	(132, 132)	0.8020	0.8016	0.0004	(130, 130)	0.8005	0.8040	− 0.0035
		0.3	(− 0.3, 0.3)	(61, 61)	0.8053	0.8056	− 0.0003	(58, 58)	0.7742	0.8001	− 0.0259
		0.4	(− 0.4, 0.4)	(36, 36)	0.8122	0.8071	0.0051	(34, 34)	0.7781	0.8082	− 0.0301
		0.2	(− 0.2, 0.2)	(98, 196)	0.8099	0.8049	0.0050	(96, 192)	0.7982	0.8047	− 0.0065
		0.3	(− 0.3, 0.3)	(45, 90)	0.8064	0.8041	0.0023	(43, 86)	0.7826	0.8016	− 0.0190
		0.4	(− 0.4, 0.4)	(27, 54)	0.8170	0.8122	0.0048	(25, 50)	0.7717	0.8077	− 0.0360

Table 3 Computed sample size, estimated power, and simulated power for the Anderson and Hauck test of simple effect when $X_F = 50$, Type I error $\alpha = 0.05$ and nominal power $1 - \beta = 0.80$

Parameter	Δ_M	Δ	(Δ_L, Δ_U)	Exact approach				Approximate method			
				(N_1, N_2)	Simulated power	Estimated power	Error	(N_1, N_2)	Simulated power	Estimated power	Error
$\mu_D = -11.5161$	-11	4	$(-15, -7)$	(76, 76)	0.8097	0.8019	0.0078	(75, 75)	0.7993	0.8034	-0.0041
		5	$(-16, -6)$	(48, 48)	0.8097	0.8006	0.0091	(47, 47)	0.7946	0.8034	-0.0088
		6	$(-17, -5)$	(34, 144)	0.8145	0.8083	0.0062	(33, 33)	0.7931	0.8103	-0.0172
		4	$(-15, -7)$	(57, 114)	0.8128	0.8037	0.0091	(56, 112)	0.8019	0.8066	-0.0047
		5	$(-16, -6)$	(36, 72)	0.8024	0.8023	0.0001	(35, 70)	0.7926	0.8032	-0.0106
		6	$(-17, -5)$	(26, 52)	0.8212	0.8191	0.0021	(25, 50)	0.8021	0.8178	-0.0157

powerful equivalence tests always have a rejection region with reasonably controlled significance level.

Within the context of linear regressions, one and two regression lines represent two major scenarios of regression slope appraisal research. Accordingly, the TOST has been applied to assess whether the linear trend is practically negligible in ecological and environmental issues. In view of the potential limitation of TOST, this study presents extended Anderson and Hauck procedures for equivalence assessment in linear regression analysis. Specifically, equivalence tests are proposed for evaluating the linear trend and mean response of a single regression line, and the trend effect and simple effect between two regression lines. The hypotheses are constructed with asymmetric equivalence bounds and therefore, can be readily applied to all equivalence problems about regression slopes and mean responses.

Moreover, to enhance the usefulness of the suggested procedures, the advanced issues of power and sample size calculations are also investigated. The proposed power and sample size procedures are derived under the random regression framework and have the distinct features to account for the imbedded uncertainty of predictor variables. It is essential to note that the recommended approaches involve statistical evaluations and iterative algorithms not currently available in statistical package. A full set of computer programs are developed for implementing the suggested equivalence tests and sample size determinations. These research findings expand the conceptual understanding and theoretical development of Anderson and Hauck procedure for equivalence assessments in linear regression analysis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42519-024-00382-7.

Funding Open Access funding enabled and organized by National Yang Ming Chiao Tung University. This work was supported by a grant from the Ministry of Science and Technology (MOST-111–2410-H-A49-034-MY3).

Availability of Data and Materials The data are presented in the article.

Declarations

Conflict of interest The author declares that he has no competing interests.

References

1. Cashen LH, Geiger SW (2004) Statistical power and the testing of null hypotheses: a review of contemporary management research and recommendations for future studies. Organ Res Methods 7:151–167. https://doi.org/10.1177/1094428104263676
2. Cortina JM, Folger RG (1998) When is it acceptable to accept a null hypothesis: no way, Jose? Organ Res Methods 1:334–350. https://doi.org/10.1177/109442819813004
3. Edwards JR, Berry JW (2010) The presence of something or the absence of nothing: increasing theoretical precision in management research. Organ Res Methods 13:668–689. https://doi.org/10.1177/1094428110380467
4. Frick RW (1995) Accepting the null hypothesis. Mem Cognit 23:132–138. https://doi.org/10.3758/bf03210562
5. Rogers JL, Howard KI, Vessey JT (1993) Using significance tests to evaluate equivalence between two experimental groups. Psychol Bull 113:553–565. https://doi.org/10.1037/0033-2909.113.3.553
6. Seaman MA, Serlin RC (1998) Equivalence confidence intervals for two-group comparisons of means. Psychol Methods 3:403–411. https://doi.org/10.1037/1082-989x.3.4.403
7. Stanton JM (2021) Evaluating equivalence and confirming the null in the organizational sciences. Organ Res Methods 24:491–512. https://doi.org/10.1177/1094428120921934
8. Stegner BL, Bostrom AG, Greenfield TK (1996) Equivalence testing for use in psychological and service research: an introduction with examples. Eval Program Plann 19:193–198. https://doi.org/10.1016/0149-7189(96)00011-0
9. Steiger JH (2004) Beyond the *F* test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. Psychol Methods 9:164–182. https://doi.org/10.1037/1082-989x.9.2.164
10. Berger RL, Hsu JC (1996) Bioequivalence trials, intersection-union tests and equivalence confidence sets (with discussion). Stat Sci 11:283–319. https://doi.org/10.1214/ss/1032280304
11. Meyners M (2012) Equivalence tests-a review. Food Qual Prefer 26:231–245. https://doi.org/10.1016/j.foodqual.2012.05.003
12. Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15:657–680. https://doi.org/10.1007/bf01068419
13. Schuirmann DL (1981) On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. Biometrics 37:617
14. Westlake WJ (1981) Response to T.B.L. Kirkwood: bioequivalence testing–a need to rethink. Biometrics 37:589–594
15. Anderson S, Hauck WW (1983) A new procedure for testing equivalence in comparative bioavailability and other clinical trials. Communications in Statistics-Theory and Methods 12:2663–2692. https://doi.org/10.1080/03610928308828634
16. Hauck WW, Anderson S (1984) A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. J Pharmacokinet Biopharm 12:83–91. https://doi.org/10.1007/bf01063612
17. Dixon PM, Pechmann JHK (2005) A statistical test to show negligible trend. Ecology 86:1751–1756. https://doi.org/10.1890/04-1343
18. Schmidt BR, Meyer AH (2008) On the analysis of monitoring data: testing for no trend in population size. J Nat Conserv 16:157–163. https://doi.org/10.1016/j.jnc.2008.05.001
19. Counsell A, Cribbie RA (2015) Equivalence tests for comparing correlation and regression coefficients. Br J Math Stat Psychol 68:292–309. https://doi.org/10.1111/bmsp.12045
20. Johnson NL, Kotz S, Balakrishnan N (1995) Continuous univariate distributions, vol 2. Wiley, New York
21. Huitema B (2011) The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies, vol 608. Wiley, New York, NY
22. Shieh G (2017) On tests of treatment-covariate interactions: An illustration of appropriate power and sample size calculations. PLoS ONE 12:e0177682. https://doi.org/10.1371/journal.pone.0177682
23. Shieh G (2020) Power analysis and sample size planning in ANCOVA designs. Psychometrika 85:101–120. https://doi.org/10.1007/s11336-019-09692-3
24. Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied linear statistical models, 5th edn. McGraw Hill, New York, NY

25. Rencher AC, Schaalje GB (2007) Linear models in statistics, 2nd edn. Wiley, Hoboken, NJ
26. Johnson PO, Neyman J (1936) Tests of certain linear hypotheses and their application to some educational problems. Stat Res Mem 1:57–93
27. Potthoff RF (1964) On the Johnson-Neyman technique and some extensions thereof. Psychometrika 29:241–256. https://doi.org/10.1007/bf02289721
28. Hunka S (1995) Identifying regions of significance in ANCOVA problems having non-homogeneous regressions. Br J Math Stat Psychol 48:161–188. https://doi.org/10.1111/j.2044-8317.1995.tb01056.x
29. Rogosa D (1980) Comparing nonparallel regression lines. Psychol Bull 88:307–321. https://doi.org/10.1037/0033-2909.88.2.307
30. Spiller SA, Fitzsimons GJ, Lynch JG Jr, McClelland GH (2013) Spotlights, floodlights, and the magic number zero: simple effects tests in moderated regression. J Mark Res 50:277–288. https://doi.org/10.1509/jmr.12.0420