

RESEARCH ARTICLE

Open Access

Sample size calculations for model validation in linear regression analysis



Show-Li Jan¹ and Gwonen Shieh^{2*}

Abstract

Background: Linear regression analysis is a widely used statistical technique in practical applications. For planning and appraising validation studies of simple linear regression, an approximate sample size formula has been proposed for the joint test of intercept and slope coefficients.

Methods: The purpose of this article is to reveal the potential drawback of the existing approximation and to provide an alternative and exact solution of power and sample size calculations for model validation in linear regression analysis.

Results: A fetal weight example is included to illustrate the underlying discrepancy between the exact and approximate methods. Moreover, extensive numerical assessments were conducted to examine the relative performance of the two distinct procedures.

Conclusions: The results show that the exact approach has a distinct advantage over the current method with greater accuracy and high robustness.

Keywords: Linear regression, Model validation, Power, Sample size, Stochastic predictor

Background

Regression analysis is the most commonly applied statistical method of all scientific fields. The extensive utility incurs continuous investigations to give various interpretations, extensions, and computing algorithms for the development and formulation of empirical models. General guidelines and fundamental principles on regression analysis have been well documented in the standard texts of Cohen et al. [1], Kutner et al. [2], and Montgomery, Peck, and Vining [3], among others. Among the methodological issues and statistical implications of regression analysis, model adequacy and validity represent two vital aspects for justifying the usefulness of the underlying regression model. In the process of model selection, residual analysis and diagnostic checking are employed to identify influential observations, leverage, outliers, multicollinearity, and other lack of fit problems. Alternatively, model validation refers to the plausibility and generalizability of the regression function in terms of the stability and suitability of the regression coefficients.

In particular, it is emphasized in Kutner et al. ([2], Section 9.6), Montgomery, Peck, and Vining ([3], Section 11.2), and Snee [4] that there are three approaches to assessing the validity of regression models: (1) comparison of model predictions and coefficients with physical theory, prior experience, theoretical models, and other simulation results; (2) collection of new data to check model predictions; and (3) data splitting in which reservation of a portion of the available data is used to obtain an independent measure of the model prediction accuracy. Essentially, the fundamental utilities between model selection and model validation should be properly recognized and distinguished because a refined model that fits the data does not necessarily guarantee prediction accuracy. Further details and related issues can be found in the importance texts of Kutner et al. [2] and Montgomery, Peck, and Vining [3] and the references therein.

The present article focuses on the validation process of linear regression analysis for comparison with postulated or acclaimed models. In linear regression, the focus is often concerned with the existence and magnitude of the slope coefficients. However, the quality of estimation

* Correspondence: gwshieh@mail.nctu.edu.tw

²Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30010, Republic of China

Full list of author information is available at the end of the article



and prediction in associating the response variable with the predictor variables is determined by the closely intertwined intercept and slope coefficients. It is of practical importance to conduct a joint test of intercept and slope coefficients in order to verify the compatibility with established or theoretical formulations. For example, Maddahi et al. [5] compared left ventricular myocardial weights of dogs by nuclear magnetic resonance imaging with actual measurements for different methods using simple linear regression analysis. The results were tested, both individually and simultaneously, whether the intercept was different from zero and the slope was different from unity. Also, Rose and McCallum [6] proposed a simple regression formula for estimating the logarithm of fetal weight with the sum of the ultrasound measurements of biparietal diameter, mean abdominal diameter, and femur length. Note that the birth weights differ among ethnic groups, cohort characteristics, and time periods. Thus, it is of considerable interest for related research to validate or compare the magnitudes of intercept and slope coefficients in their formulation.

The importance and implications of statistical power analysis in research studies are well addressed in Cohen [7], Kramer and Blasey [8], Murphy, Myros, and Wolach [9], and Ryan [10], among others. In the context of multiple regression and correlation, the distinct notions of fixed and random regression settings were emphasized and explicated in power and sample size calculations by Gatsonis and Sampson [11], Mendoza and Stafford [12], Sampson [13], and Shieh [14–16]. On the other hand, Kelley [17], Krishnamoorthy and Xia [18], and Shieh [19] discussed sample size determinations for constructing precise confidence intervals of strength of association. It is noteworthy that analysis of covariance (ANCOVA) models involving both categorical and continuous predictors incur different hypothesis testing procedures. Accordingly, they require unique power procedures as discussed in Shieh [20] and Tang [21], among others.

For the purposes of planning research designs and validating model formulation, a sample size procedure was presented in Colosimo et al. [22]. The presented formula has a computationally appealing expression and maintains reasonable accuracy in their simulation study. However, the particular method involves a convenient substitution of fixed mean parameter for random predictor variables. Their illustrations were not detailed enough to address the extent and impact of such simplification in sample size computations. Consequently, the adequacy of the sample size procedure described in Colosimo et al. [22] requires further clarification and no research to date has examined its properties under different situations.

The statistical inferences for the regression coefficients are based on the conditional distribution of the

continuous predictors. However, unlike the fixed factor configurations and treatment levels in analysis of variance (ANOVA) and other experimental designs, the continuous measurements of the predictor variables in regression studies are typically available only after the data has been collected. For advance planning research design, the distribution and power functions of the test procedure need to be appraised over possible values of the predictors. Thus, it is important to recognize the stochastic nature of the predictor variables. The fundamental differences between fixed and random models have been explicated in Binkley and Abbot [23], Cramer and Appelbaum [24], Sampson [13], and Shaffer [25]. Despite the complexity associated with the unconditional properties of the test procedure, the inferential procedures are the same under both fixed and random formulations. Hence, the usual rejection rule and critical value remain unchanged. The distinction between the two modeling approaches becomes critical for power analysis and sample size planning.

The joint test of intercept and slope coefficients in linear regression is more involved than the individual tests of intercept or slope parameters. A general linear hypothesis setting is required to perform the simultaneous test of both intercept and slope coefficients as shown in Rencher and Schaalje ([26], Section 8.4.2). However, it is essential to emphasize that they did not address the corresponding power and sample size issues. In view of the limited results in current literature, this article aims to present power and sample size procedure for the joint test of intercept and slope coefficients with specific recognition of the stochastic features of predictor variables. First, exact power function and sample size procedure for detecting intercept and slope differences of simple linear regression are derived under random modeling framework assuming predictor variables have independent and identical normal distribution. Then, the technical presentation is extended to the general context of multiple linear regression. Then, a numerical example of model validation is employed to demonstrate the essential discrepancy between the exact and approximate methods. The accuracy and robustness of the contending methods are appraised through simulation studies under a wide range of model configurations with normal and non-normal predictors.

Methods

Simple linear regression

Consider the simple linear regression model for associating the response variable Y with the predictor variable X :

$$Y_i = \beta_I + X_i\beta_S + \varepsilon_i, \tag{1}$$

where Y_i is the observed value of the response variable Y ; X_i is the recorded value of the continuous predictor X ; β_I and β_S are unknown intercept and slope parameters; and ε_i are *iid* $N(0, \sigma^2)$ random errors for $i = 1, \dots, N$. To examine the existence and magnitude of the intercept and slope coefficients $\{\beta_I, \beta_S\}$, the statistical inferences are based on the least squares estimators $\hat{\beta}_I$ and $\hat{\beta}_S$, where $\hat{\beta}_I = \bar{Y} - \bar{X}\hat{\beta}_S$, $\hat{\beta}_S = SSXY/SSX$, $\bar{Y} = \sum_{i=1}^N Y_i/N$, $\bar{X} = \sum_{i=1}^N X_i/N$, $SSXY = \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$, and $SSX = \sum_{i=1}^N (X_i - \bar{X})^2$. It follows from the standard results in Rencher and Schaalje ([26], Section 7.6.3) that the estimators $\{\hat{\beta}_I, \hat{\beta}_S\}$ have the bivariate normal distribution

$$\hat{\boldsymbol{\beta}} \sim N_2(\boldsymbol{\beta}, \sigma^2 \mathbf{W}_X), \tag{2}$$

where

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_I \\ \hat{\beta}_S \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_I \\ \beta_S \end{bmatrix}, \quad \mathbf{W}_X = \begin{bmatrix} W_{X11} & W_{X12} \\ W_{X21} & W_{X22} \end{bmatrix},$$

$W_{X11} = 1/N + \bar{X}^2/SSX$, $W_{X12} = W_{X21} = -\bar{X}/SSX$, and $W_{X22} = 1/SSX$. The subscript X of \mathbf{W}_X emphasizes the elements $\{W_{X11}, W_{X12}, W_{X21}, W_{X22}\}$ of the variance and covariance matrix are functions of the predictor variables. Also, $\hat{\sigma}^2 = SSE/v$ is the usual unbiased estimator of σ^2 , where $SSE = SSY - SSXY^2/SSX$ is the error sum of squares, $SSY = \sum_{i=1}^N (Y_i - \bar{Y})^2$, and $v = N - 2$. Note that the least squares estimators $\hat{\beta}_I$ and $\hat{\beta}_S$ are independent of $\hat{\sigma}^2$.

A joint test of the intercept and slope coefficients can be conducted with the hypothesis

$$H_0 : \begin{bmatrix} \beta_I \\ \beta_S \end{bmatrix} = \begin{bmatrix} \beta_{I0} \\ \beta_{S0} \end{bmatrix} \text{ versus } H_1 : \begin{bmatrix} \beta_I \\ \beta_S \end{bmatrix} \neq \begin{bmatrix} \beta_{I0} \\ \beta_{S0} \end{bmatrix}. \tag{3}$$

Following the model assumption in Eq. 1, the likelihood ratio statistic for the joint test of intercept and slope is

$$F_J = \frac{(\hat{\boldsymbol{\beta}}_D^T \mathbf{W}_X^{-1} \hat{\boldsymbol{\beta}}_D)/2}{\hat{\sigma}^2}, \tag{4}$$

where $\hat{\boldsymbol{\beta}}_D = [\hat{\beta}_{ID}, \hat{\beta}_{SD}]^T$, $\hat{\beta}_{ID} = \hat{\beta}_I - \beta_{I0}$, and $\hat{\beta}_{SD} = \hat{\beta}_S - \beta_{S0}$. Under the null hypothesis, it can be shown that

$$F_J \sim F(2, v), \tag{5}$$

where $F(2, v)$ is an F distribution with 2 and v degrees of freedom. Hence, H_0 is rejected at the significance level α if

$$F_J > F_{2, v, \alpha}, \tag{6}$$

where $F_{2, v, \alpha}$ is the upper $(100-\alpha)$ th percentile of the $F(2, v)$ distribution. In general, the joint test statistic F_J has the nonnull distribution for the given values of \bar{X} and SSX :

$$F_J | [\bar{X}, SSX] \sim F(2, v, \Delta_J) \tag{7}$$

where

$$\Delta_J = \frac{\{N(\beta_{ID} + \bar{X}\beta_{SD})^2 + \beta_{SD}^2 SSX\}}{\sigma^2}. \tag{8}$$

Hence, the noncentral F distribution $F(c, v, \Delta_J)$ is a function of the predictor values $\{X_i, i = 1, \dots, N\}$ only through the summary statistics \bar{X} and SSX .

The joint test of the intercept and slope coefficients given in Eq. 3 can be viewed as a special case of the general linear hypothesis considered in Rencher and Schaalje ([26], Section 8.4.2). However, two important aspects of this study should be pointed out. First, unlike the current consideration, the associated F test and related statistical properties in Rencher and Schaalje [26] are presented under the standard settings with fixed predictor values. Second, they did not address the power and sample size issues under random modeling formulations. Accordingly, their fundamental results are extended here to accommodate the predictor features in power and sample size calculations for the validation of simple linear regression models.

The statistical inferences about the regression coefficients are based on the conditional distribution of the continuous variables $\{X_i, i = 1, \dots, N\}$. Therefore, the resulting analysis would be specific to the observed values of the predictors. It is clear that, before conducting a research study, the actual values of predictors are not available beforehand just as the major responses. In view of the stochastic nature of the summary statistics \bar{X} and SSX , it is essential to recognize and assess the distribution of the test statistic over possible values of the predictors. To demonstrate the impact of the predictor features on power and sample size calculations, the normality setting is commonly employed to provide a convenient basis for analytical derivation and empirical examination of random predictors as in Gatsonis and Sampson [11], Sampson [13], and Shieh [14]. However, it is important to note that the power and sample size calculations of Gatsonis and Sampson [11], Sampson [13], and Shieh [14, 15] for detecting slope coefficients in multiple regression analysis are not applicable for assessing differences in intercept and slope coefficients considered here.

Specifically, the continuous predictor variables $\{X_i, i = 1, \dots, N\}$ are assumed to have independent and identical normal distribution $N(\mu_X, \sigma_X^2)$. With the normal assumption,

it can be readily established that $\bar{X} \sim N(\mu_X, \sigma_X^2/N)$ and $K = SSX/\sigma_X^2 \sim \chi^2(\kappa)$ where $\kappa = N - 1$. Thus, the noncentrality Δ_J in Eq. 8 can be expressed as

$$\Delta_J = \frac{\{N(a + bZ)^2 + dK\}}{\sigma^2}, \tag{9}$$

where $a = \beta_{ID} + \mu_X\beta_{SD}$, $b = (d/N)^{1/2}$, $d = \beta_{SD}^2\sigma_X^2$, and $Z = (\bar{X} - \mu_X)/(\sigma_X^2/N)^{1/2} \sim N(0, 1)$. As a consequence, the F_J statistic has the two-stage distribution

$$F_J | [K, Z] \sim F(2, v, \Delta_J), \quad K \sim \chi^2(\kappa),$$

and $Z \sim N(0, 1)$.

(10)

Note that the two random variables K and Z are independent. Moreover, the corresponding power function for the simultaneous test can be formulated as

$$\Psi_J = E_K E_Z [P\{F(2, v, \Delta_J) > F_{2,v,\alpha}\}], \tag{11}$$

where the expectations E_K and E_Z are taken with respect to the distributions of K and Z , respectively.

Alternatively, Colosimo et al. ([22], Section 3.2) described a simple and naive method to obtain an unconditional distribution of F_J . They substituted the sample values of the predictor variables in the noncentrality Δ_J with the corresponding expected value $E[X_i] = \mu_X$ for $i = 1, \dots, N$. Thus, the distribution of F_J is approximated by a noncentral F distribution:

$$F_J \sim F(2, v, \Delta_C), \tag{12}$$

where $\Delta_C = (Na^2)/\sigma^2$. The suggested power function of Colosimo et al. [22] for the joint test of intercept and slope coefficients is

$$\Psi_C = P\{F(2, v, \Delta_C) > F_{2,v,\alpha}\}. \tag{13}$$

It is vital to note that the approximate power function Ψ_C only involves a noncentral F distribution, whereas the normal predictor distributions lead to the exact and more complex power formula Ψ_J that consists of a joint chi-square and normal mixture of noncentral F distributions. Evidently, the power function Ψ_C is relatively simpler to compute than the exact formula Ψ_J . But the approximate nature of Ψ_C does not involve all of the predictor features in power computations.

It follows from large sample theory that Z and K/N converge to 0 and 1, respectively. Hence, the sample-size-adjusted noncentrality quantity Δ_J/N approaches Δ_J^* as the sample size N increases to infinity, where

$$\Delta_J^* = \frac{(\beta_{ID} + \mu_X\beta_{SD})^2 + \beta_{SD}^2\sigma_X^2}{\sigma^2}. \tag{14}$$

Hence, Δ_J^* provides a convenient measurement of effect size for the joint appraisal of intercept and slope coefficients. It can be immediately seen from the noncentrality term of the approximate power function Ψ_C that $\Delta_C^* = \Delta_C/N = (\beta_{ID} + \mu_X\beta_{SD})^2/\sigma^2 < \Delta_J^*$ with the exceptions that $\beta_{SD} = 0$ and/or $\sigma_X^2 = 0$. Consequently, the estimated power Ψ_C is generally less than that of Ψ_J even for large sample sizes when all other configurations remain constant. It is shown later that while the computation is more involved for the complex power function Ψ_J , the exact approach has a clear advantage over the approximate procedure in accurate power calculations. For advance planning of a research design, the presented power formulas can be employed to calculate the sample size N needed to attain the specified power $1 - \beta$ for the chosen significance level α , null values $\{\beta_{I0}, \beta_{S0}\}$, coefficient parameters $\{\beta_I, \beta_S\}$, variance component σ^2 , and predictor mean and variance $\{\mu_X, \sigma_X^2\}$. It usually involves an incremental search with a small initial value to find the optimal solution for achieving the desired power performance.

Multiple linear regression

The power and sample size calculations for the general scenario of multiple linear regression with more than one predictor are discussed next. Consider the multiple linear regression model with response variable Y_i and p predictor variables (X_{i1}, \dots, X_{ip}) for $i = 1, \dots, N$:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{15}$$

where $\mathbf{Y} = (Y_1, \dots, Y_N)^T$ is an $N \times 1$ vector with Y_i being the observed measurement of the i th subject; $\mathbf{X} = (\mathbf{1}_N, \mathbf{X}_S)$ with $\mathbf{1}_N$ is the $N \times 1$ vector of all 1's, $\mathbf{X}_S = (\mathbf{X}_{S1}, \dots, \mathbf{X}_{SN})^T$ is an $N \times p$ matrix, $\mathbf{X}_{Si} = (X_{i1}, \dots, X_{ip})^T$, X_{i1}, \dots, X_{ip} are the observed values of the p predictor variables of the i th subject; $\boldsymbol{\beta} = (\beta_I, \boldsymbol{\beta}_S^T)^T$ is a $(p + 1) \times 1$ vector with $\boldsymbol{\beta}_S = (\beta_1, \dots, \beta_p)^T$ and $\beta_I, \beta_1, \dots, \beta_p$ are unknown coefficient parameters; and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$ is an $N \times 1$ vector with ε_i are *iid* $N(0, \sigma^2)$ random variables.

For the joint test of intercept and slope coefficients in terms of

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\theta} \text{ versus } H_1 : \boldsymbol{\beta} \neq \boldsymbol{\theta}, \tag{16}$$

it can be shown from Rencher and Schaalje ([6], Section 8.4.2) that the test statistic is

$$F_{MJ} = \frac{\left\{ (\hat{\boldsymbol{\beta}} - \boldsymbol{\theta})^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\theta}) \right\} / (p + 1)}{\hat{\sigma}^2}, \tag{17}$$

where $\hat{\sigma}^2 = SSE/v$ is the usual unbiased estimator of σ^2 . Under the null hypothesis, F_{MJ} has an F distribution with $p + 1$ and v degrees of freedom

$$F_{MJ} \sim F(p + 1, v) \tag{18}$$

The joint test can be conducted by reject H_0 at the significance level α if $F_{MJ} > F_{(p+1), v, \alpha}$. In general, F_{MJ} has the nonnull distribution for the given values of \mathbf{X}_S :

$$F_{MJ} \sim F(p + 1, v, \Delta_{MJ}), \tag{19}$$

where $F(p + 1, v, \Delta_{MJ})$ is a noncentral F distribution with $p + 1$ and v degrees of freedom and noncentrality parameter Δ_{MJ} with

$$\Delta_{MJ} = \frac{\{(\boldsymbol{\beta} - \boldsymbol{\theta})^T (\mathbf{X}^T \mathbf{X}) (\boldsymbol{\beta} - \boldsymbol{\theta})\}}{\sigma^2}. \tag{20}$$

It is essential to emphasize that the inferences in Rencher and Schaalje [26] are concerned mainly with the slope coefficients $\boldsymbol{\beta}_S$. As noted in the context of simple linear regression, the fundamental results concerning fixed predictor values are extended here to power and sample size calculations for the validation of linear regression models under random predictor settings.

In view of the random nature of the predictor variables, the continuous predictor variables $\{\mathbf{X}_{Si}, i = 1, \dots, N\}$ are assumed to have independent multinormal distributions $N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$. With the multinormal assumptions, it can be readily established that $\bar{\mathbf{X}}_S = \sum_{i=1}^N \mathbf{X}_{Si}/N \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X/N)$ and $\mathbf{A} = \sum_{i=1}^N (\mathbf{X}_{Si} - \bar{\mathbf{X}}_S)(\mathbf{X}_{Si} - \bar{\mathbf{X}}_S)^T \sim W_p(\kappa, \boldsymbol{\Sigma}_X)$, where $W_p(\kappa, \boldsymbol{\Sigma}_X)$ is a Wishart distribution with κ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}_X$, and $\kappa = N - 1$. Thus, the noncentrality Δ_{MJ} can be rewritten as

$$\Delta_{MJ} = \frac{\{N(\boldsymbol{\beta}_{ID} + \boldsymbol{\beta}_{SD}^T \bar{\mathbf{X}}_S)^2 + \boldsymbol{\beta}_{SD}^T \mathbf{A} \boldsymbol{\beta}_{SD}\}}{\sigma^2}, \tag{21}$$

where $\boldsymbol{\beta}_{ID} = \boldsymbol{\beta}_I - \boldsymbol{\theta}_I$ and $\boldsymbol{\beta}_{SD} = \boldsymbol{\beta}_S - \boldsymbol{\theta}_S$. Using the prescribed distributions of $\bar{\mathbf{X}}_S$ and \mathbf{A} , it can be shown that $\boldsymbol{\beta}_{ID} + \boldsymbol{\beta}_{SD}^T \bar{\mathbf{X}}_S = a + bZ \sim N(a, b^2)$, $Z \sim N(0, 1)$, and $K = \boldsymbol{\beta}_{SD}^T \mathbf{A} \boldsymbol{\beta}_{SD}/d \sim \chi^2(\kappa)$, where $a = \boldsymbol{\beta}_{ID} + \boldsymbol{\beta}_{SD}^T \boldsymbol{\mu}_X$, $b = (d/N)^{1/2}$, and $d = \boldsymbol{\beta}_{SD}^T \boldsymbol{\Sigma}_X \boldsymbol{\beta}_{SD}$. Note that the two random variables K and Z are independent. It is conceptually simple and computationally convenient to subsume the stochastic features of $\bar{\mathbf{X}}_S$ and \mathbf{A} in terms of Z and K . Accordingly, the noncentrality quantity Δ_J is formulated as

$$\Delta_{MJ} = \frac{\{N(a + bZ)^2 + dK\}}{\sigma^2}. \tag{22}$$

Thus, under the multinormal predictor assumptions, the F_{MJ} statistic has the two-stage distribution

$$F_{MJ} | [K, Z] \sim F(p + 1, v, \Delta_{MJ}), K \sim \chi^2(\kappa), \text{ and } Z \sim N(0, 1). \tag{23}$$

The corresponding power function for the simultaneous test can be termed as

$$\Psi_{MJ} = E_K E_Z [P\{F(p + 1, v, \Delta_{MJ}) > F_{(p+1), v, \alpha}\}], \tag{24}$$

where the expectations E_K and E_Z are taken with respect to the distribution of K and Z , respectively. Evidently, when $p = 1$, the test statistic F_{MJ} and power function Ψ_{MJ} reduce to the simplified formulas of F_M and Ψ_J given in Eqs. 4 and 11, respectively.

Results

An illustration

To demonstrate the prescribed power and sample size procedures, the simplified formula for estimating fetal weight in Rose and McCallum [6] is used as a benchmark for validation. Although there are several different methods for estimating the fetal weight, it was demonstrated in Anderson et al. [27] that the simple linear regression formula of Rose and McCallum [6] compares favorably with other techniques. Based on the ultrasound examinations conducted in the Stanford University Hospital labor and delivery suite between January 1981 and March 1984, they presented a useful formula for predicting the natural logarithm of birth weight with the sum of head, abdomen, and limb ultrasound measurements as given by the equation: $\ln(\text{BW}) = 4.198 + 0.143 \cdot X$, where $X = \text{biparietal diameter} + \text{mean abdominal diameter} + \text{femur length}$ (in centimeters). The average birth weight of their study population was 2275 g with a range of 490–5300 g. The detailed comparisons and related discussions of viable equations for estimating fetal weight can be found in Anderson et al. [27] and the references therein.

Conceivably, there are underlying differences in fetal weight between different ethnic origins, cohort groups, and time periods. To validate the simple formula for a target population, it requires a detailed scheme to determine the necessary sample size so that the conducted study has a decent assurance in detecting the potential discrepancy. For illustration, the intercept and slope coefficients are set as $\beta_I = 4.1$ and $\beta_S = 0.15$, respectively. The error component is selected to be $\sigma^2 = 0.095$. The characteristics of the ultrasound measurements are represented by the mean $\mu_X = 24.2$ and variance $\sigma_X^2 = 6$. Note that these configurations assure that the expected fetal weight of the designated population $E[\text{BW}] = E[\exp(4.1 + 0.15 \cdot X + \epsilon)] = 2275.52$ coincides the average magnitude of birth weighs reported in Rose and

McCallum [6]. To test the hypothesis of $H_0: (\beta_I, \beta_S) = (4.198, 0.143)$ versus $H_1: (\beta_I, \beta_S) \neq (4.198, 0.143)$ with the significance level $\alpha = 0.05$, numerical computations showed that the sample sizes of $N_E = 173$ and 227 are required for the exact approach to attain the target power of 0.8 and 0.9 , respectively. Because of the sample sizes need to be integer values in practice, the attained power is slightly greater than the nominal power level. In these two cases, the achieved powers of the two sample sizes are $\Psi_J = 0.8001$ and 0.9010 , respectively. These results were computed with the supplementary algorithms presented in Additional files 1 and 2. For ease of application, the prescribed configurations are incorporated in the user specification sections of the SAS/IML programs.

On the other hand, the matching sample sizes computed with the approximate method of Colosimo et al. [22] are $N_C = 183$ and 239 with the attained powers of $\Psi_C = 0.8010$ and 0.9002 , respectively. Therefore, the simple method of Colosimo et al. [22] clearly requires $183 - 173 = 10$ and $239 - 227 = 12$ more babies than the exact formula to satisfy the nominal power performance. Actually, the exact power function gives the values $\Psi_J = 0.8236$ and 0.9161 with the sample sizes 183 and 239 , respectively. Hence, the resulting power differences between the two magnitudes of sample size are $0.8236 - 0.8001 = 0.0235$ and $0.9161 - 0.9010 = 0.0151$. To enhance the illustration, the computed sample size, estimated power, and difference for the exact and approximate procedures are summarized in Table 1. The sample size and power calculations show that the approximate power function Ψ_C tends to underestimate powers because the simplification of noncentrality parameter in the noncentral F distribution. Correspondingly, the approximate method of Colosimo et al. [22] often overestimates the required sample sizes for validation analysis. It is essential to note that adopting a small sample size will cause a study that has insufficient power to demonstrate model difference. In this case of Colosimo et al. [22], their method may lead to an over-sized study that wastes time, money, and other resources. More importantly, the hypothesis tests of validation studies are over-rejected and yield erroneous conclusions. It is of both practical usefulness and theoretical concern to further assess the intrinsic implications of the two distinct

procedures for other settings. Detailed empirical studies are described next to evaluate and compare their accuracy under a wide variety of model configurations.

Numerical comparisons

In view of the potential discrepancy between the exact and approximate procedures, numerical investigations of power and sample size calculations were conducted under a wide range of model configurations in two studies. The first assessment focuses on the situations with normal predictor variables, while the second study concerns the robustness of the two methods under several prominent situations of non-normal predictors.

Normal predictors

For ease of comparison, the model settings in Colosimo et al. [22] are considered and expanded to reveal the distinct behavior of the contending procedures. Specifically, the null and alternative hypotheses are

$$H_0 : \begin{bmatrix} \beta_I \\ \beta_S \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ versus } H_1 : \begin{bmatrix} \beta_I \\ \beta_S \end{bmatrix} \neq \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

where $\{\beta_I, \beta_S\} = \{d, 1 + d\}$ and $\{\beta_{ID}, \beta_{SD}\} = \{d, d\}$ with $d = 0.3, 0.4,$ and 0.5 . Note that these coefficient settings are equivalent to those with $\{\beta_I, \beta_S\} = \{\beta_{I0} + d, \beta_{S0} + d\}$ because they lead to the same differences $\{\beta_{ID}, \beta_{SD}\} = \{d, d\}$ and the resulting power functions remain identical. The error component is fixed as $\sigma^2 = 1$ and the predictors X are assumed to have normal distributions with mean $\mu_X = \{0, 0.5, 1\}$ and variance $\sigma_X^2 = \{0.5, 1, 2\}$. Overall these considerations result in a total of 27 different combined settings. These combinations of model configurations were chosen to represent the possible characteristics that are likely to be encountered in actual applications and also to maintain a reasonable range for the magnitudes of sample size without making unrealistic assessments.

Throughout this empirical investigation, the significance level and nominal power are fixed as $\alpha = 0.05$ and $1 - \beta = 0.90$, respectively. With the prescribed specifications, the required sample sizes are computed for the exact procedure with the power function Ψ_J . The computed sample sizes of the nine combined predictor mean and variance patterns are summarized in Table 2, Table S1 and Table S2 for the coefficient difference $d = 0.3, 0.4,$ and 0.5 , respectively. As suggested by a referee, Tables S1 and S2 are presented in Additional files 3 and 4, respectively. In order to evaluate the accuracy of power calculations, the estimated power of the exact and approximate procedures are also presented. Note that the attained values of the exact approach are marginally larger than the nominal level 0.90 . In contrast, the estimated powers of the approximation of Colosimo et al. [22] are all less than 0.90 and the difference is quite

Table 1 Computed sample size, estimated power, and difference for the exact and approximate procedures with $\{\beta_I, \beta_S\} = \{4.1, 0.15\}$, $\{\beta_{I0}, \beta_{S0}\} = \{4.198, 0.143\}$, $\sigma^2 = 0.095$, $\mu_X = 24.2$, $\sigma_X^2 = 6$, and Type I error $\alpha = 0.05$

Nominal power	Exact approach		Approximate method		Difference	
	N	Estimated power	N	Estimated power	N	Power
0.80	173	0.8001	183	0.8236	10	0.0235
0.90	227	0.9010	239	0.9161	12	0.0151

Table 2 Computed sample size, estimated power, and simulated power for Normal predictors with $\{\beta_I, \beta_S\} = \{0.3, 1.3\}$, $\{\beta_{I0}, \beta_{S0}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$

μ_X	σ_X^2	N	Simulated power	Exact approach		Approximate method	
				Simulated power	Error	Simulated power	Error
0	0.5	99	0.9049	0.9025	-0.0024	0.7524	-0.1525
	1	76	0.8997	0.9030	0.0033	0.6257	-0.2740
	2	53	0.9058	0.9050	-0.0008	0.4602	-0.4456
0.5	0.5	56	0.9029	0.9055	0.0026	0.8430	-0.0599
	1	48	0.8993	0.9024	0.0031	0.7756	-0.1237
	2	38	0.8997	0.9006	0.0009	0.6604	-0.2393
1	0.5	35	0.9015	0.9013	-0.0002	0.8682	-0.0333
	1	33	0.9075	0.9089	0.0014	0.8445	-0.0630
	2	28	0.8993	0.9016	0.0023	0.7689	-0.1304

substantial in some cases. Then, Monte Carlo simulation studies of 10,000 iterations were performed to compute the simulated power for the designated sample sizes and parameter configurations. For each replicate, N predictor values were generated from the designated normal distribution $N(\mu_X, \sigma_X^2)$. The resulting values of normal predictor, intercept and slope coefficients $\{\beta_I, \beta_S\}$, and error variance σ^2 , in turn, determine the configurations for producing N normal outcomes of the simple linear regression model defined in Eq. 1. Next, the test statistic F_J was computed and the simulated power was the proportion of the 10,000 replicates whose test statistics F_J exceed the corresponding critical value $F_{2, v, 0.05}$. The adequacy of the two sample size procedures is determined by the error between the estimate power and the simulated power. The simulated power and error are also summarized in Table 2, Table S1 and Table S2 for all twenty-seven design schemes.

It can be seen from these results that the discrepancy between the estimated power and the simulated power is considerably small for the proposed exact technique for all model configurations considered here. Specifically, the resulting errors of the 27 designs are all within the small range of -0.0087 to 0.0056. On the other hand, the estimated powers of the approximate method are constantly smaller than the simulated powers. The outcomes show a clear pattern that absolute error decreases with coefficient difference d and predictor mean μ_X , and increases with predictor variance σ_X^2 , when all other configurations are held constant. Notably, the associated absolute errors can be as large as 0.4456, 0.4295, and 0.4183 when $\mu_X = 0$ and $\sigma_X^2 = 2$ for $d = 0.3, 0.4,$ and 0.5 in Table 2, Table S1, Table S2, respectively. It should be noted that most of the sample sizes reported in the empirical examination of Colosimo et al. [22] (Table 1) are rather large and impractical. This may explain why the performance of the approximate formula was acceptable in their study. In fact, some of their cases with smaller

sample sizes also showed the same phenomenon that the simple method leads to an underestimate of power level and an overestimated sample size required to achieve the nominal power. Essentially, the simplicity of the approximate formula does come with a huge price in terms of inaccurate power and sample size calculations.

Non-normal predictors

To address the sensitivity issues of the two techniques, power and sample size calculations were also conducted for the regression models with non-normal predictors. For illustration, the model settings in Table 2 with $\{\beta_{ID}, \beta_{SD}\} = \{0.3, 0.3\}$ are modified by assuming the predictors have four different sets of distributions: Exponential(1), Gamma(2, 1), Laplace(1), and Uniform(0, 1). For ease of comparison, the designated distributions were linearly transformed to have mean μ_X and variance σ_X^2 as reported in the previous study. Hence, the computed sample sizes associated with the exact procedure and estimated powers of the two methods remain identical for the four non-normal distributions. The simulated powers were obtained with the Monte Carlo simulation studies of 10,000 iterations under the selected model configurations and non-normal predictor distributions. Similar to the numerical assessments in the preceding study, the computed sample sizes, simulated powers, estimated powers, and associated errors of the two competing procedures are presented in Tables S3-S6 of Additional files 5, 6, 7, 8 for the four types of non-normal predictors, respectively.

Regarding the robustness properties of the two procedures, the results in suggest that the performance of the exact approach is slightly affected by the non-normal covariate settings. The high skewness and kurtosis of the Exponential distribution apparently has a more prominent impact on the normal-based power function than the other three cases of Gamma, Laplace, and Uniform distributions. Note that the approximate

method only depends on the mean values of the predictors and is presumably less sensitive to the variation of predictor distributions. However, the accuracy marginally improved in some cases, but generally maintains almost the same performance as in the normal setting presented in Table 2. In short, the sensitivity and robustness of the suggested exact technique depends on the level of how badly predictor distributions depart from normality structure. On the other hand, the performance assessments show that the exact procedure still give acceptable results even in the situations with non-normal predictors considered here. More importantly, these empirical evidences reveal that the exact approach is relatively more reliable and accurate than the approximate method to be recommended as a trustworthy technique for power and sample calculations.

Discussion

In practice, a research study requires adequate statistical power and sufficient sample size to detect scientifically credible effects. Although multiple linear regression is a well-recognized statistical tool, the corresponding power and sample size problem for model validation has not been adequately examined in the literature. To enhance the usefulness of the joint test of intercept and slope coefficients in linear regression analysis, this article presents theoretical discussions and computational algorithms for power and sample size calculations under the random modeling framework. The stochastic nature of predictor variables is taken into account by assuming that they have an independent and identical normal distribution. In contrast, the existing method of Colosimo et al. [22] adopted a direct replacement of mean values for the predictor variables. Consequently, the proposed exact approach has the prominent advantage of accommodating the complete distributional features of normal predictors whereas the simple approximation of Colosimo et al. [22] only includes the mean parameters of the predictor variables.

Conclusions

The presented analytic derivations and empirical results indicate that the approximate formula of Colosimo et al. [22] generally does not give accurate power and sample size calculations. According to the overall accuracy and robustness, the exact approach clearly outperforms the approximate methods as a useful tool in planning validation study. Although the numerical illustration only involves a predictor variable, it embodies the underlying principle and critical feature of linear regression that can be useful in conducting similar evaluations for the more general framework of multiple linear regression.

Additional files

Additional file 1: SAS/IML program for computing the power for the joint test of intercept and slope coefficients. (PDF 60 kb)

Additional file 2: SAS/IML program for computing the sample size for the joint test of intercept and slope coefficients. (PDF 60 kb)

Additional file 3: Table S1. Computed sample size, estimated power, and simulated power for Normal predictors with $\{\beta_1, \beta_2\} = \{0.4, 1.4\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 95 kb)

Additional file 4: Table S2. Computed sample size, estimated power, and simulated power for Normal predictors with $\{\beta_1, \beta_2\} = \{0.5, 1.5\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 96 kb)

Additional file 5: Table S3. Computed sample size, estimated power, and simulated power for transformed Exponential predictors with $\{\beta_1, \beta_2\} = \{0.3, 1.3\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 95 kb)

Additional file 6: Table S4. Computed sample size, estimated power, and simulated power for transformed Gamma predictors with $\{\beta_1, \beta_2\} = \{0.3, 1.3\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 95 kb)

Additional file 7: Table S5. Computed sample size, estimated power, and simulated power for transformed Laplace predictors with $\{\beta_1, \beta_2\} = \{0.3, 1.3\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 95 kb)

Additional file 8: Table S6. Computed sample size, estimated power, and simulated power for transformed Uniform predictors with $\{\beta_1, \beta_2\} = \{0.3, 1.3\}$, $\{\beta_{10}, \beta_{20}\} = \{0, 1\}$, $\sigma^2 = 1$, Type I error $\alpha = 0.05$, and nominal power $1 - \beta = 0.90$. (PDF 96 kb)

Abbreviations

ANCOVA: Analysis of covariance; ANOVA: Analysis of variance

Acknowledgements

The authors would like to thank the editor and two reviewers for their constructive comments that led to an improved article.

Funding

No funding.

Availability of data and materials

The summary statistics are available from the following article: [6].

Authors' contributions

SLJ conceived of the study, and participated in the development of theory and helped to draft the manuscript. GS carried out the numerical computations, participated in the empirical analysis and drafted the manuscript. Both authors read and approved the final manuscript.

Authors' information

SLJ is a professor of Applied Mathematics, Chung Yuan Christian University, Taoyuan, Taiwan 32023. GS is a professor of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30010.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Applied Mathematics, Chung Yuan Christian University, Taoyuan, Taiwan 32023, Republic of China. ²Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30010, Republic of China.

Received: 31 August 2018 Accepted: 26 February 2019

Published online: 12 March 2019

References

- Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Mahwah: Erlbaum; 2003.
- Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. 5th ed. New York: McGraw Hill; 2005.
- Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 5th ed. Hoboken: Wiley; 2012.
- Snee RD. Validation of regression models: methods and examples. *Technometrics*. 1977;19:415–28.
- Maddahi J, Crues J, Berman DS, et al. Noninvasive quantification of left ventricular myocardial mass by gated proton nuclear magnetic resonance imaging. *J Am Coll Cardiol*. 1987;10:682–92.
- Rose BI, McCallum WD. A simplified method for estimating fetal weight using ultrasound measurements. *Obstet Gynecol*. 1987;69:671–4.
- Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Erlbaum; 1988.
- Kraemer HC, Blasey C. How many subjects?: Statistical power analysis in research. 2nd ed. Los Angeles: Sage; 2015.
- Murphy KR, Myers B, Wolach A. Statistical power analysis: a simple and general model for traditional and modern hypothesis tests. 4th ed. New York: Routledge; 2014.
- Ryan TP. Sample size determination and power. Hoboken: Wiley; 2013.
- Gatsonis C, Sampson AR. Multiple correlation: exact power and sample size calculations. *Psychol Bull*. 1989;106:516–24.
- Mendoza JL, Stafford KL. Confidence interval, power calculation, and sample size estimation for the squared multiple correlation coefficient under the fixed and random regression models: a computer program and useful standard tables. *Educ Psychol Meas*. 2001;61:650–67.
- Sampson AR. A tale of two regressions. *J Am Stat Assoc*. 1974;69:682–9.
- Shieh G. Exact interval estimation, power calculation and sample size determination in normal correlation analysis. *Psychometrika*. 2006;71:529–40.
- Shieh G. A unified approach to power calculation and sample size determination for random regression models. *Psychometrika*. 2007;72:347–60.
- Shieh G. Exact analysis of squared cross-validity coefficient in predictive regression models. *Multivar Behav Res*. 2009;44:82–105.
- Kelley K. Sample size planning for the squared multiple correlation coefficient: accuracy in parameter estimation via narrow confidence intervals. *Multivar Behav Res*. 2008;43:524–55.
- Krishnamoorthy K, Xia Y. Sample size calculation for estimating or testing a nonzero squared multiple correlation coefficient. *Multivar Behav Res*. 2008;43:382–410.
- Shieh G. Sample size requirements for interval estimation of the strength of association effect sizes in multiple regression analysis. *Psicothema*. 2013;25:402–7.
- Shieh G. Power and sample size calculations for contrast analysis in ANCOVA. *Multivar Behav Res*. 2017;52:1–11.
- Tang Y. Exact and approximate power and sample size calculations for analysis of covariance in randomized clinical trials with or without stratification. *Stat Biopharm Res*. 2018;10:274–86.
- Colosimo EA, Cruz FR, Miranda JLO, et al. Sample size calculation for method validation using linear regression. *J Stat Comput Simul*. 2007;77:505–16.
- Binkley JK, Abbot PC. The fixed X assumption in econometrics: can the textbooks be trusted? *Am Stat*. 1987;41:206–14.
- Cramer EM, Appelbaum MI. The validity of polynomial regression in the random regression model. *Rev Educ Res*. 1978;48:511–5.
- Shaffer JP. The Gauss-Markov theorem and random regressors. *Am Stat*. 1991;45:269–73.
- Rencher AC, Schaafje GB. Linear models in statistics. 2nd ed. Hoboken: Wiley; 2007.
- Anderson NG, Jolley IJ, Wells JE. Sonographic estimation of fetal weight: comparison of bias, precision and consistency using 12 different formulae. *Ultrasound Obstet Gynecol*. 2007;30:173–9.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

