Journal of Quality Measurement and Analysis *JQMA* 22(2) 2026, 157-174
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2202.2026.09

AN EZEKIEL'S ADJUSTED R-SQUARED-BASED DECISION RULE FOR REGRESSION HYPOTHESIS TESTING

(*Peraturan Keputusan Berasaskan R-Kuasa Dua Terlaras Ezekiel bagi Pengujian Hipotesis Regresi*)

SET FOONG NG*

ABSTRACT

Prior studies have reported that a minimal bias-variance tradeoff for Ezekiel's adjusted R-squared provides a more accurate representation of a model's explanatory power. Motivated by the advantages of Ezekiel's adjusted R-squared, this study proposes its integration into the framework of null hypothesis significance testing for regression models. The primary objective is to develop a new decision rule for the F-test in regression analysis based on Ezekiel's adjusted R-squared and to evaluate the influence of various parameters (the number of explanatory variables, sample size and alpha value) on the critical value in the new decision rule. This new adjusted R-squared based decision rule is useful because it provides researchers a way to gauge the practical significance of a model at the same time testing statistical significance. In this study, an analysis using linear regression models based on forty-eight cases revealed that low values of adjusted R-squared were observed in statistically significant regression models. The results of this study are consistent with prior research, in which conclusions derived from hypothesis testing are regarded as uncertain and have been subject to criticism. In conclusion, the hypothesis testing procedure, including its application in linear regression analysis, need not be dismissed outright. Rather, it should be regarded as one of several screening tools to be interpreted in conjunction with other relevant factors and domain-specific considerations.

Keywords: linear regression model; Ezekiel's adjusted R-squared; hypothesis testing; regression analysis

ABSTRAK

Kajian-kajian terdahulu telah melaporkan bahawa kompromi kecondongan-varian yang minimum bagi R-kuasa dua terlaras Ezekiel memberikan gambaran yang lebih tepat terhadap keupayaan penjelasan model. Didorong oleh kelebihan R-kuasa dua terlaras Ezekiel, kajian ini mencadangkan penyepaduannya ke dalam rangka kerja ujian kepentingan hipotesis nol untuk model regresi. Objektif utama adalah untuk membangunkan peraturan keputusan baharu bagi ujian-F dalam analisis regresi berdasarkan R-kuasa dua terlaras Ezekiel dan untuk menilai pengaruh pelbagai parameter (bilangan pembolehubah penjelasan, saiz sampel dan nilai alfa) pada nilai kritikal dalam peraturan keputusan baharu. Peraturan keputusan baharu berasaskan R-kuasa dua terlaras ini berguna kerana ia menyediakan penyelidik dengan satu kaedah untuk menilai kepentingan praktikal sesuatu model pada masa yang sama dengan pengujian kepentingan statistik. Dalam kajian ini, analisis menggunakan model regresi linear berdasarkan empat puluh lapan kes menunjukkan bahawa nilai R-kuasa dua terlaras yang rendah diperoleh dalam model regresi yang signifikan secara statistik. Keputusan kajian ini bersesuaian dengan kajian terdahulu di mana kesimpulan yang diperoleh daripada ujian hipotesis dianggap sebagai tidak pasti dan telah tertakluk kepada kritikan. Kesimpulannya, prosedur ujian hipotesis, termasuk penggunaannya dalam analisis regresi linear, tidak perlu diketepikan secara langsung. Sebaliknya, ia harus dianggap sebagai salah satu alat penyaringan untuk ditafsirkan bersama dengan faktor lain yang berkaitan dan pertimbangan khusus mengikut domain.

Kata kunci: model regresi linear; R-kuasa dua terlaras Ezekiel; pengujian hipotesis; analisis regresi

1. Introduction

The linear regression model is a widely used statistical method for describing the relationship between a dependent variable and a set of explanatory variables. It has been extensively applied across various fields of research, including the prediction of economic growth in financial and economic research (Ebiwonjumi *et al.* 2023), the assessment of insulin absorption in clinical application (Arpaia *et al.* 2022), the forecasting of air pollution levels within environmental research (Mozumder *et al.* 2013; Hsu *et al.* 2025) and various applications in experimental studies (Meng *et al.* 2021; Ng 2025).

In linear regression model construction, the squared multiple correlation coefficient is used to represent the true proportion of variance in the dependent variable that is explained by the set of explanatory variables. This coefficient is denoted by ρ^2.

The R-squared is commonly used as an estimator of the squared multiple correlation coefficient. It is typically denoted as R^2. R-squared gives a measure of how well the values of dependent variables are explained by the regression model constructed from the sample. However, it is well recognized that the R-squared tends to increase as additional explanatory variables are included in the regression model. In such cases, R-Squared may overestimate the proportion of variance explained by the model (Karch 2020). In other words, R-Squared can overestimate the squared multiple correlation coefficient, making it a positively biased estimator of the squared multiple correlation coefficient.

As an alternative to R-squared, several formulations of adjusted R-squared, such as Smith formula, Wherry formula and Ezekiel formula, have been proposed in the literature as shrinkage estimators of the squared multiple correlation coefficient (Wherry 1931; Huberty & Mourad 1980; Krus & Fuller 1982; Ayabe 1985; Kennedy 1988; Kromrey & Hines 1995). The adjusted R-squared is designed to correct the value of R-Squared by accounting for both sample size and model complexity in the formula (McClave *et al.* 2001). Model complexity is influenced by the number of explanatory variables included in the regression analysis.

Studies by Hittner (2020) and Raju *et al.* (1999) reported that Ezekiel's adjusted R-squared exhibits the lowest bias when compared to other estimators (including Smith and Wherry formula)in which its performance in terms of root mean squared error (RMSE) is comparable to other measures. This indicates a minimal bias-variance tradeoff for Ezekiel's adjusted R-squared. Consequently, Ezekiel's adjusted R-squared provides a more accurate representation of the model's explanatory power. Currently, many statistical software packages commonly employ the Ezekiel formula to compute and report the adjusted R-squared value (Yin & Fan 2001).

The F-test in regression analysis is a null hypothesis significance testing commonly used to evaluate the significance of the regression coefficients in linear regression models (Wang & Cui 2013). Specifically, it assesses whether the model accounts for a significant proportion of the variance in the dependent variable. The test statistic in the F-test is calculated based on the explained sum of squares (ESS) and the residual sum of squares (RSS). According to the classical decision rule, if the test statistic in terms of ESS and RSS exceeds the critical value, the null hypothesis in the F-test is rejected, indicating that the regression model is statistically significant. Motivated by the advantages of Ezekiel's adjusted R-squared in regression analysis (Raju *et al.* 1999; Hittner 2020), this study proposes its integration into the framework of null hypothesis significance testing for regression models. The objective of this paper is to develop a new decision rule for F-test in regression analysis based on Ezekiel's adjusted R-squared and to evaluate the influence of various parameters (number of explanatory variables, sample size and alpha value) on the critical value in the new decision rule. The findings of this study provide an insight into the corresponding adjusted R-squared value for a statistically significant regression model based on the significance level in the classical decision rule for F-test. This new adjusted R-squared based decision rule is useful because it provides researchers a way to gauge the practical significance of a model at the same time testing statistical significance.

The rest of the paper is organized as follows. Section 2 outlines the formulation of the squared multiple correlation coefficient for linear regression model, Ezekiel's adjusted R-squared and the classical decision rule for F-test in regression analysis. Section 3 details the derivation of new decision rule for F-test in regression analysis. The evaluation of the influence of various parameters on the critical value in the new decision rule for F-test is presented in Section 4. Section 5 concludes the work.

2. Preliminaries

The formulation of the squared multiple correlation coefficient for linear regression model, residual sum of squares, explained sum of squares, total sum of squares, R-squared, Ezekiel's adjusted R-squared and the classical decision rule for F-test in regression analysis are presented in this section.

2.1. *The squared multiple correlation coefficient*

Linear regression model that describes the relationship of a dependent variable y and a set of k explanatory variables $x_1, x_2, \cdots, x_k$ is represented by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + e, \tag{1}$$

where β_j are the regression coefficients and e is the residual error.

Assuming the dependent variable is normally distributed with a constant variance σ_T^2, where σ_T^2 is called total variance. Suppose the residual error follows a normal distribution such that $e \sim N(0, \sigma^2)$, where σ^2 is called error variance.

In many cases, the main objective of regression models is to predict the values of dependent variable based on the observed data. The squared multiple correlation coefficient is used to represent the true proportion of variance explained by the explanatory variables (Karch 2020). The squared multiple correlation coefficient ρ^2 is represented by

$$\rho^2 = 1 - \frac{\sigma^2}{\sigma_T^2}, \tag{2}$$

where σ^2 is the error variance and σ_T^2 is the total variance.

2.2. *The residual sum of squares, explained sum of squares and total sum of squares*

In regression model, residuals are the deviations between the observed values and the predicted values of the dependent variable. Residual sum of squares (RSS) is the sum of squares of residuals. RSS is a measure of discrepancy between the estimated regression model and the observed data. Small value of RSS shows a good fit of the regression model to the observed data. In linear regression model $y = \beta_0 + \beta_1 x + e$ with one explanatory variable x, RSS is represented by

$$RSS = \sum_{i=1}^{n} \left(y_i - y_i \right)^2 , \tag{3}$$

where y_i is the observed value of the dependent variable, y_i is the predicted value of y_i and n is sample size of the observed data.

The explained sum of squares (ESS), also known as regression sum of squares, is the sum of squares of the difference between the predicted value and the mean value of the dependent variable. ESS describes how well a regression model fits the data. ESS is represented by

$$ESS = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 , \tag{4}$$

where y_i is the predicted value of y_i and $\overline{y}$ is the mean value of y_i.

The total sum of squares (TSS) measures how much variation there is in the observed values and the mean value of the dependent variable. TSS is represented by

$$TSS = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 , \tag{5}$$

where y_i is the observed value of the dependent variable and $\overline{y}$ is the mean value of y_i.

Total sum of squares (TSS) is equal to the sum of the explained sum of squares (ESS) and the residual sum of squares (RSS). In simple linear regression model $y = \beta_0 + \beta_1 x + e$, the relationship between these sum of squares is represented by

$$TSS = ESS + RSS , \tag{6}$$

$$\sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 + \sum_{i=1}^{n} \left(y_i - y_i \right)^2 . \tag{7}$$

In multiple linear regression model $\mathbf{Y} = \mathbf{X\beta} + \mathbf{e}$ with k explanatory variables $x_1, x_2, \cdots, x_k$, the ordinary least squares estimator (OLSE) for the parameter $\boldsymbol{\beta}$ is given as follows.

$$\boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{Y} . \tag{8}$$

The predicted dependent variable is obtained as follows.

$$\begin{aligned} \mathbf{Y} &= \mathbf{X\beta} \\ &= \mathbf{X}\left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{Y} \end{aligned} \tag{9}$$

RSS, ESS and TSS in multiple linear regression model are given as follows.

$$\begin{aligned} RSS &= \left(\mathbf{Y}-\mathbf{Y}\right)'\left(\mathbf{Y}-\mathbf{Y}\right) \\ &= \left(\mathbf{Y}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\right)'\left(\mathbf{Y}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\right) \\ &= \mathbf{Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \end{aligned} \tag{10}$$

$$\begin{aligned} ESS &= \left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) \\ &= \mathbf{Y}'\mathbf{Y}-2\mathbf{Y}'\overline{\mathbf{Y}}+\overline{\mathbf{Y}}'\overline{\mathbf{Y}} \\ &= \mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}-2\mathbf{Y}'\overline{\mathbf{Y}}+\overline{\mathbf{Y}}'\overline{\mathbf{Y}} \end{aligned} \tag{11}$$

$$\begin{aligned} TSS &= \left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) \\ &= \mathbf{Y}'\mathbf{Y}-2\mathbf{Y}'\overline{\mathbf{Y}}+\overline{\mathbf{Y}}'\overline{\mathbf{Y}}, \end{aligned} \tag{12}$$

where $\overline{\mathbf{Y}}$ is the mean of observed values in the dependent variable vector $\mathbf{Y}$.

It is noted that the relationship between the sum of squares in multiple linear regression model is also represented by $TSS = ESS + RSS$. The equation can be easily proved by noting that $\mathbf{Y}'\overline{\mathbf{Y}} = \mathbf{Y}'\overline{\mathbf{Y}}$ as the sum of residual is zero, i.e. $\sum_{i=1}^{n}\left(y_i - y_i\right) = 0$.

To summarize, the TSS measures how much variation there is in the observed values. The ESS measures how much variation there is in the predicted values. The RSS measures the variation in the error between observed values and predicted values.

2.3. *The R-squared*

Researchers have proposed various formulae to estimate the squared multiple correlation coefficient ρ^2. R-squared is one of the commonly used estimator for ρ^2. R-squared quantifies the proportion of total variation of dependent variable that can be explained by the explanatory variables in the sample.

R-squared is obtained from the equation of ρ^2, in which the error variance σ^2 is replaced with $\sigma^2 = \frac{RSS}{n}$ while the total variance σ_T^2 is replaced with $\sigma_T^2 = \frac{TSS}{n}$. Hence, the formula of R-squared is given as follows (Walpole *et al.* 2007).

$$\begin{aligned} R^2 &= 1-\frac{\sigma^2}{\sigma_T^2} \\ &= 1-\frac{RSS/n}{TSS/n} \\ &= 1-\frac{RSS}{TSS}\ , \end{aligned} \tag{13}$$

where RSS is the residual sum of squares and TSS is the total sum of squares.

2.4. *The Ezekiel's adjusted R-squared*

Another estimator for the squared multiple correlation coefficient ρ^2 is called adjusted R-squared. It is a statistical measure that is a modified form of R-squared. It measures the total proportion of variance of dependent variable that is explained by the regression model, accounting for the number of explanatory variables.

Various formulae are used to compute adjusted R-squared, $\bar{R}^2$. Some variations of the formula of $\bar{R}^2$ are given by the Ezekiel formula $\bar{R}_E^2$, the Wherry formula $\bar{R}_W^2$, and the Smith formula $\bar{R}_S^2$.

The Ezekiel formula is obtained from modifying the equation of ρ^2, in which the error variance σ^2 is replaced with $\hat{\sigma}^2 = \dfrac{RSS}{n-k-1}$ while the total variance σ_T^2 is replaced with $\hat{\sigma}_T^2 = \dfrac{TSS}{n-1}$. Hence, the Ezekiel formula of adjusted R-squared is represented by

$$\begin{aligned}\bar{R}_E^2 &= 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_T^2} \\ &= 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)} \\ &= 1 - \frac{n-1}{n-k-1}\left(\frac{RSS}{TSS}\right), \end{aligned} \tag{14}$$

where n is the sample size and k is the number of explanatory variables in the regression models.

From Eq. (13), we obtain $\dfrac{RSS}{TSS} = 1 - R^2$. Therefore, the Ezekiel formula can also be expressed as follows.

$$\bar{R}_E^2 = 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right). \tag{15}$$

Ezekiel formula was given different names in literatures. It was named as Wherry-1 formula by studies in Krus and Fuller (1982) and Ayabe (1985). Besides, Ezekiel formula was also named as Cohen formula by Kennedy (1988). Ezekiel formula is used by many statistical packages, such as SPSS and SAS/STAT, to report the value of adjusted R-squared (Yin & Fan 2001).

The Wherry formula was first proposed by Wherry (1931). Some studies have adopted this formula in the estimation of the squared multiple correlation coefficient (Huberty & Mourad 1980; Kromrey & Hines 1995). The Wherry formula of adjusted R-squared is represented by

$$\bar{R}_W^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right). \tag{16}$$

The Smith formula was first presented by M. J. B. Ezekial in the year 1928 during the meeting of American Mathematical Society held at Chicago (Wherry, 1931). The Smith formula of adjusted R-squared is represented by

$$R_S^2 = 1 - \frac{n}{n-k}\left(1 - R^2\right). \tag{17}$$

The results in the study by Raju *et al.* (1999) concluded that the Ezekiel formula produced the most accurate estimate, followed by the Smith formula and Wherry formula.

2.5. *The classical decision rule for F-test in regression analysis*

The linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + e$ is given in Eq. (1). Hypothesis testing is widely used in regression analysis to evaluate the overall significance of the regression model. The test statistic of the hypothesis testing follows an F distribution (Milton & Arnold 2004). Hence, the hypothesis testing is commonly referred to the F-test in regression analysis. Essentially, the F-test checks whether the explanatory variables are useful predictors of the dependent variable.

The hypothesis testing in F-test compares the model's fit to a model with no explanatory variables, that is a model with only constant term (β_0). The null hypothesis of F-test in regression analysis, H_0, states that all regression coefficients, excluding the intercept (β_0), have values of zero. In other words, H_0 in F-test suggests that none of the explanatory variables has a significant effect on the dependent variable.

The alternative hypothesis of F-test in regression analysis, H_1, states that at least one regression coefficient (excluding the intercept) has value that is not equal to zero. In other words, H_1 in F-test suggests that at least one explanatory variable has a significant effect on the dependent variable. The null and alternative hypothesis of F-test in regression analysis are given as follows.

$$\begin{cases} H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0 \\ H_1 : \text{at least one } \beta_j \neq 0, \quad j = 1, 2, \cdots, k \end{cases} \tag{18}$$

The test statistic of the F-test (F_{TS}) is a ratio of variances (McClave *et al.* 2001). F_{TS} is defined as the ratio of regression mean square to the residuals mean square. Regression mean square is equal to the explained variability divided by the degrees of freedom for the regression model ($v_1 = k$). Residuals mean square is equal to the unexplained variability divided by the degrees of freedom for residual ($v_2 = n - k - 1$). The formulation of F_{TS} is given in Eq. (19).

$$\begin{aligned} F_{TS} &= \frac{ESS / v_1}{RSS / v_2} \\ &= \frac{ESS / k}{RSS / (n-k-1)}. \end{aligned} \tag{19}$$

The value of test statistic F_{TS} tends to be large when the variance explained by the regression model is much higher than the unexplained variance. To determine when the test statistic is large enough that one can confidently reject the null hypothesis and conclude that the model is useful for predicting dependent variable, we would compare the value of F_{TS} with a critical value F_{CV}. The critical value is a threshold such that if the test statistic exceeds it (i.e. $F_{TS} > F_{CV}$), the F-test indicates that the null hypothesis is not true. Thus, H_0 should be rejected.

The critical value for the F-test at a significance level of α is given by

$$F_{CV} = F_{\alpha, \nu_1, \nu_2}, \tag{20}$$

where $\nu_1 = k$ and $\nu_2 = n-k-1$.

The F-test in regression analysis is a right-tailed test such that the critical value divides the F-distribution into a rejection region (G_1) and a non-rejection region (G_1'). The regions are given as follows.

$$\begin{cases} G_1 = \{F_{TS} \mid F_{TS} > F_{CV}\} \\ G_1' = \{F_{TS} \mid F_{TS} \le F_{CV}\} \end{cases}, \tag{21}$$

where $F_{TS} = \dfrac{ESS/k}{RSS/(n-k-1)}$, $F_{CV} = F_{\alpha, \nu_1, \nu_2}$, $\nu_1 = k$ and $\nu_2 = n-k-1$.

The classical decision rule for the F-test in regression analysis is given as follows.

$$D_1: \text{Reject } H_0 \text{ if } F_{TS} \in G_1 \text{ such that } G_1 = \{F_{TS} \mid F_{TS} > F_{CV}\}. \tag{22}$$

3. Derivation of New Decision Rule for F-test in Regression Analysis

In this section, a new decision rule for F-test in regression analysis based on Ezekiel's adjusted R-squared is derived.

The formula of Ezekiel's adjusted R-squared $R_E^2 = 1 - \dfrac{n-1}{n-k-1}\left(\dfrac{RSS}{TSS}\right)$ in Eq. (14) is referred. Here, we present $\dfrac{RSS}{TSS}$ and $\dfrac{ESS}{TSS}$ in terms of R_E^2 in Eq. (23) and Eq. (24), respectively.

$$\frac{RSS}{TSS} = \left(1 - R_E^2\right)\left(\frac{n-k-1}{n-1}\right) \tag{23}$$

$$\begin{aligned} \frac{ESS}{TSS} &= 1 - \frac{RSS}{TSS} \\ &= 1 - \left(1 - R_E^2\right)\left(\frac{n-k-1}{n-1}\right) \end{aligned} \tag{24}$$

Theorem 3.1. *Given a linear regression model with a dependent variable and a set of k explanatory variables. The decision rule for the F-test in regression analysis based on Ezekiel's adjusted R-squared is given as follows.*

$$\text{Reject } H_0 \text{ if } R_E^2 \in G_2 \text{ such that } G_2 = \{R_E^2 \,|\, R_E^2 > (R_E^2)_{CV}\},$$

where $(R_E^2)_{CV} = \dfrac{kF_{\alpha,\nu_1,\nu_2} - k}{kF_{\alpha,\nu_1,\nu_2} + n - k - 1}$, $\nu_1 = k$ *and* $\nu_2 = n - k - 1$.

Proof. Using $F_{TS} = \dfrac{ESS/k}{RSS/(n-k-1)}$ in Eq. (19), $\dfrac{RSS}{TSS} = \left(1 - R_E^2\right)\left(\dfrac{n-k-1}{n-1}\right)$ in Eq. (23) and $\dfrac{ESS}{TSS} = 1 - \left(1 - R_E^2\right)\left(\dfrac{n-k-1}{n-1}\right)$ in Eq. (24), the classical rejection region $G_1 = \{F_{TS} \,|\, F_{TS} > F_{CV}\}$ in Eq. (21) is transformed into a new rejection region G_2 based on R_E^2. The derivation of G_2 is shown below.

$$G_1 = \{F_{TS} \,|\, F_{TS} > F_{CV}\}$$

$$G_1 = \left\{F_{TS} \,\middle|\, \frac{ESS/k}{RSS/(n-k-1)} > F_{CV}\right\}$$

$$G_1 = \left\{F_{TS} \,\middle|\, \frac{ESS/TSS}{RSS/TSS}\left(\frac{n-k-1}{k}\right) > F_{CV}\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, \left[\frac{1 - \left(1 - R_E^2\right)\left(\dfrac{n-k-1}{n-1}\right)}{\left(1 - R_E^2\right)\left(\dfrac{n-k-1}{n-1}\right)}\right]\left(\frac{n-k-1}{k}\right) > F_{CV}\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, \frac{n-1}{k\left(1 - R_E^2\right)} - \left(\frac{n-k-1}{k}\right) > F_{CV}\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, \frac{n-1}{1 - R_E^2} > kF_{CV} + n - k - 1\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, \frac{n-1}{kF_{CV} + n - k - 1} > 1 - R_E^2\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, R_E^2 > \frac{kF_{CV} - k}{kF_{CV} + n - k - 1}\right\}$$

$$G_2 = \left\{R_E^2 \,\middle|\, R_E^2 > \frac{kF_{\alpha,\nu_1,\nu_2} - k}{kF_{\alpha,\nu_1,\nu_2} + n - k - 1}\right\}$$

$$G_2 = \{R_E^2 \,|\, R_E^2 > (R_E^2)_{CV}\}, \tag{25}$$

where $(R_E^2)_{CV} = \dfrac{kF_{\alpha,\nu_1,\nu_2} - k}{kF_{\alpha,\nu_1,\nu_2} + n - k - 1}$, $\nu_1 = k$ and $\nu_2 = n - k - 1$.

It can be seen from Eq. (25) that R_E^2 can be used as the test statistic for F-test in regression analysis while $(R_E^2)_{CV}$ is the corresponding critical value for F-test. If the test statistic exceeds the critical value (i.e. $R_E^2 > (R_E^2)_{CV}$), the F-test indicates that the null hypothesis in Eq. (18) should be rejected. Hence, the decision rule for the F-test in regression analysis based on Ezekiel formula of adjusted R-squared is derived as follows.

$$D_2: \text{ Reject } H_0 \text{ if } R_E^2 \in G_2 \text{ such that } G_2 = \{R_E^2 \mid R_E^2 > (R_E^2)_{CV}\}, \quad (26)$$

where $(R_E^2)_{CV} = \dfrac{kF_{\alpha,\nu_1,\nu_2} - k}{kF_{\alpha,\nu_1,\nu_2} + n - k - 1}$, $\nu_1 = k$ and $\nu_2 = n - k - 1$.

The proof is complete. □

4. Results and Discussion

In this section, a study is performed to evaluate the influence of various parameters (number of explanatory variables, sample size and alpha value) on the critical value $(R_E^2)_{CV}$ in the new decision rule for the F-test in regression analysis. It is noted that the probability of the rejection region in the F-test in regression analysis is represented by the significance level. Here, alpha value α is used to represent the significance level. The alpha value is the threshold of deciding if the test statistic R_E^2 is large enough to reject null hypothesis in Eq. (18).

4.1. *The overview design of the study*

In this study, linear regression models based on forty-eight cases are designed to evaluate the influence of various parameters on the critical value $(R_E^2)_{CV}$ in the new decision rule for the F-test. Three parameters being studied are as follows:

(1) Number of explanatory variables in the linear regression model, k
(2) Sample size, n
(3) Alpha value in the F-test, α

For $\alpha = 0.05$, the values of $(R_E^2)_{CV}$ are evaluated based on linear regression models formed by varying $k = 1, 2, 5, 10$ and sample size $n = 20, 30, 50, 100$ through sixteen cases (Case 1 to 16). In addition, another eight cases (Case 17 to 24) are also studied by varying $k = 1, 2, 5, 10$ and large sample size $n = 200, 400$.

For $\alpha = 0.01$, the values of $(R_E^2)_{CV}$ are investigated based on sixteen linear regression models formed by varying $k = 1, 2, 5, 10$ and $n = 20, 30, 50, 100$ through sixteen cases (Case 25 to 40). For large sample size $n = 200, 400$, eight cases (Case 41 to 48) are also studied by varying $k = 1, 2, 5, 10$. All forty-eight cases are listed in Table 1.

Table 1: Summary of linear regression models based on forty-eight cases formed by varying parameters k , n , α

	$k=1$	$k=2$	$k=5$	$k=10$
$n=20$	[a]Case 1 [b]Case 25	[a]Case 5 [b]Case 29	[a]Case 9 [b]Case 33	[a]Case 13 [b]Case 37
$n=30$	[a]Case 2 [b]Case 26	[a]Case 6 [b]Case 30	[a]Case 10 [b]Case 34	[a]Case 14 [b]Case 38
$n=50$	[a]Case 3 [b]Case 27	[a]Case 7 [b]Case 31	[a]Case 11 [b]Case 35	[a]Case 15 [b]Case 39
$n=100$	[a]Case 4 [b]Case 28	[a]Case 8 [b]Case 32	[a]Case 12 [b]Case 36	[a]Case 16 [b]Case 40
$n=200$	[a]Case 17 [b]Case 41	[a]Case 19 [b]Case 43	[a]Case 21 [b]Case 45	[a]Case 23 [b]Case 47
$n=400$	[a]Case 18 [b]Case 42	[a]Case 20 [b]Case 44	[a]Case 22 [b]Case 46	[a]Case 24 [b]Case 48

[a]Cases based on $\alpha = 0.05$.
[b]Cases based on $\alpha = 0.01$.

4.2. *Evaluation of the influence of various parameters on the critical value in the new decision rule for F-test*

Table 2 summarizes the results of the critical values $(R_E^2)_{CV}$ in the new decision rule for the F-test generated from Case 1 to 16 where sample sizes $n = 20, 30, 50, 100$ and $\alpha = 0.05$ are considered. The cases for $k = 1$ represent simple linear regression model that consists of only one explanatory variable. The cases with k more than 1, (i.e. $k = 2, 5, 10$) refer to multiple linear regression models consisting of two, five and ten explanatory variables.

Table 2: Critical value $(R_E^2)_{CV}$ in the F-test for Case 1 to 16 when $\alpha = 0.05$

	k	n	$(R_E^2)_{CV}$
Case 1	1	20	0.152
Case 2	1	30	0.099
Case 3	1	50	0.058
Case 4	1	100	0.029
Case 5	2	20	0.214
Case 6	2	30	0.140
Case 7	2	50	0.082
Case 8	2	100	0.041
Case 9	5	20	0.340
Case 10	5	30	0.218
Case 11	5	50	0.127
Case 12	5	100	0.062
Case 13	10	20	0.529
Case 14	10	30	0.322
Case 15	10	50	0.181
Case 16	10	100	0.087

For large sample size $n = 200$ and $n = 400$, the critical values $(R_E^2)_{CV}$ based on $\alpha = 0.05$ are generated by another eight cases, that are Case 17 to 24. Table 3 summarizes the results of the critical values $(R_E^2)_{CV}$ for these cases.

Table 3: Critical value $(R_E^2)_{CV}$ in the F-test for Case 17 to 24 when $\alpha = 0.05$

	k	n	$(R_E^2)_{CV}$
Case 17	1	200	0.014
Case 18	1	400	0.007
Case 19	2	200	0.020
Case 20	2	400	0.010
Case 21	5	200	0.031
Case 22	5	400	0.015
Case 23	10	200	0.042
Case 24	10	400	0.021

Based on $\alpha = 0.01$, the critical values $(R_E^2)_{CV}$ in the new decision rule for the F-test are generated by Case 25 to 40 with sample size $n = 20, 30, 50, 100$. The results of the critical values $(R_E^2)_{CV}$ for the mentioned cases are presented in Table 4.

Table 4: Critical value $(R_E^2)_{CV}$ in the F-test for Case 25 to 40 when $\alpha = 0.01$

	k	n	$(R_E^2)_{CV}$
Case 25	1	20	0.277
Case 26	1	30	0.186
Case 27	1	50	0.112
Case 28	1	100	0.056
Case 29	2	20	0.350
Case 30	2	30	0.236
Case 31	2	50	0.143
Case 32	2	100	0.072
Case 33	5	20	0.493
Case 34	5	30	0.333
Case 35	5	50	0.201
Case 36	5	100	0.101
Case 37	10	20	0.691
Case 38	10	30	0.456
Case 39	10	50	0.270
Case 40	10	100	0.134

Considering large sample size $n = 200$ and $n = 400$, the critical values $(R_E^2)_{CV}$ based on $\alpha = 0.01$ are generated by another eight cases (Case 41 to 48). Table 5 presents the results of the critical values $(R_E^2)_{CV}$ for these cases.

Table 5: Critical value $(R_E^2)_{CV}$ in the F-test for Case 41 to 48 when $\alpha = 0.01$

	k	n	$(R_E^2)_{CV}$
Case 41	1	200	0.028
Case 42	1	400	0.014
Case 43	2	200	0.036
Case 44	2	400	0.018
Case 45	5	200	0.050
Case 46	5	400	0.025
Case 47	10	200	0.066
Case 48	10	400	0.033

The graphs of critical values $(R_E^2)_{CV}$ in F-test based on different number of explanatory variables k are presented in Figure 1 for sample size $n = 20$ (Figure 1a), $n = 30$ (Figure 1b), $n = 50$ (Figure 1c) and $n = 100$ (Figure 1d).

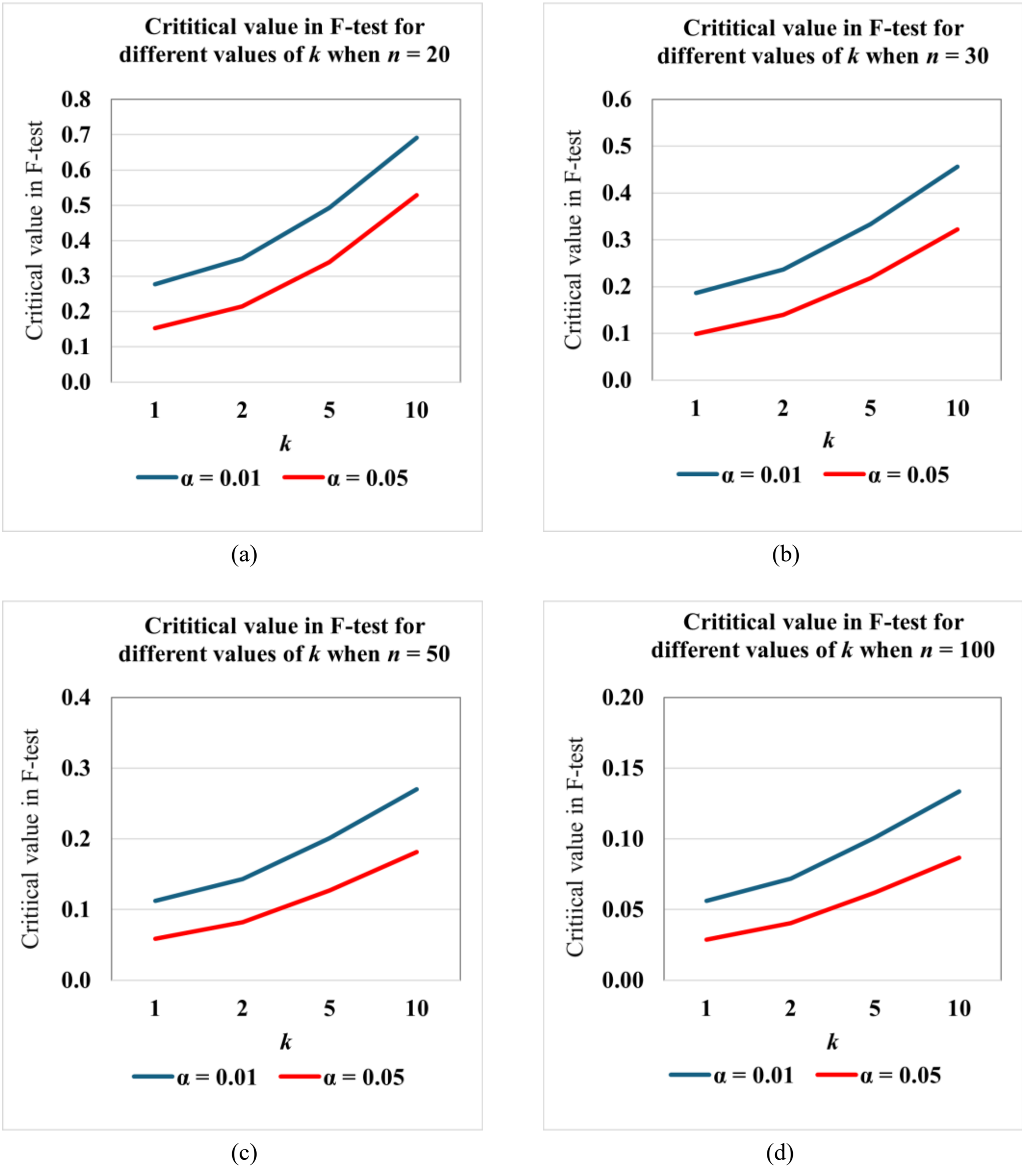
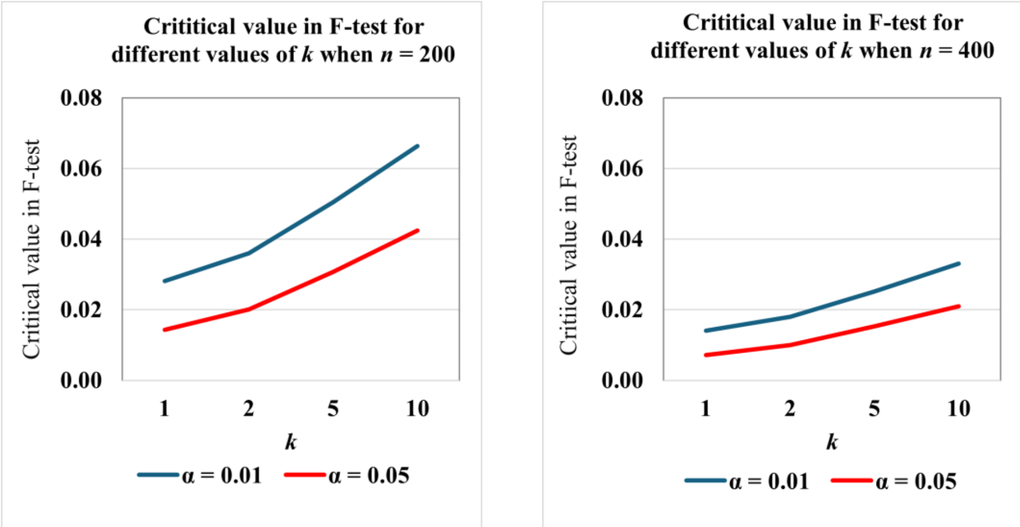


Figure 1: Critical value $(R_E^2)_{CV}$ in F-test based on different number of explanatory variables k and different sample size (a) $n = 20$, (b) $n = 30$, (c) $n = 50$ and (d) $n = 100$

For large sample size, the graphs of critical values $(R_E^2)_{CV}$ in the new decision rule for the F-test versus k are presented in Figure 2 for cases when $n = 200$ (Figure 2a) and $n = 400$ (Figure 2b).



(a) (b)

Figure 2: Critical value $(R_E^2)_{CV}$ in F-test based on different number of explanatory variables k and different sample size (a) $n = 200$ and (b) $n = 400$

4.2.1. *The influence of alpha value on the critical value*

It is observed that the values of $(R_E^2)_{CV}$ based on $\alpha = 0.01$ are higher than those based on $\alpha = 0.05$ across all cases presented in Figure 1 and Figure 2. This is due to the rejection region in F-test based on $\alpha = 0.01$ is smaller than the rejection region based on $\alpha = 0.05$. Furthermore, the F-test in regression analysis is a right-tailed test. Therefore, a larger value in the critical value $(R_E^2)_{CV}$ based on $\alpha = 0.01$ forming a smaller rejection region compared to those cases based on $\alpha = 0.05$. An example taken from the cases as shown in Figure 1b has illustrated this scenario. The rejection region (right-tailed) formed by the critical value $(R_E^2)_{CV} = 0.456$ at the significance level $\alpha = 0.01$ is smaller than the rejection region (right-tailed) formed by the critical value $(R_E^2)_{CV} = 0.322$ at the significance level $\alpha = 0.05$ for $k = 10$.

4.2.2. *The influence of the number of explanatory variables on the critical value*

It is also observed that the values of $(R_E^2)_{CV}$ tends to increase when the model complexity increases across all cases presented in Figure 1 and Figure 2. When the number of explanatory variables k increases, the model complexity increases. An example taken from the cases with sample size $n = 20$ as shown in Figure 1a has illustrated this scenario. In this paper, Theorem 3.1 states that the null hypothesis in the F-test should be rejected if the test statistic R_E^2 exceeds the critical value $(R_E^2)_{CV}$. Referring to the point $\left(k, (R_E^2)_{CV}\right) = (1, 0.277)$ in Figure 1a, it shows that H_0 in the F-test is rejected at the significance level $\alpha = 0.01$ when

$R_E^2 > 0.277$ in the case of simple linear regression model. On the other hand, a linear regression model consisting of two explanatory variables resulting in $\left(k, (R_E^2)_{CV}\right) = (2, 0.350)$ implies that H_0 is rejected at the significance level $\alpha = 0.01$ when $R_E^2 > 0.350$. An increase of the critical value $(R_E^2)_{CV}$ in a more complex model shows that a larger value of adjusted R-squared is needed for a more complex regression model to be statistically significant.

4.2.3. *The influence of sample size on the critical value*

Across all cases shown in Figure 1 and Figure 2, it is observed that the values of $(R_E^2)_{CV}$ are decreasing when sample size increases. The points taken from Figure 1a to Figure 1d, Figure 2a and Figure 2b illustrate this scenario. Based on $\alpha = 0.05$ and $k = 5$, the points $\left(n, (R_E^2)_{CV}\right)$ with different sample sizes are compared, i.e. $(20, 0.340)$, $(30, 0.218)$, $(50, 0.127)$, $(100, 0.062)$, $(200, 0.031)$ and $(400, 0.015)$. It shows that H_0 in the F-test is rejected at the significance level $\alpha = 0.05$ when $R_E^2 > 0.340$ for a model with sample size of 20. However, H_0 in the F-test is rejected when $R_E^2 > 0.031$ for $n = 200$ and $R_E^2 > 0.015$ for $n = 400$. It can be observed that the values of $(R_E^2)_{CV}$ tend to be relatively higher in regression models with smaller sample sizes ($n = 20, 30, 50, 100$) compared to those with larger sample sizes ($n = 200, 400$).

Taking an example of $\left(n, (R_E^2)_{CV}\right) = (400, 0.015)$ with $k = 5$ to explain the statistical significance in the regression model based on the F-test. It can be anticipated that the F-test in regression analysis may indicate that a regression model with an adjusted R-squared slightly above 0.015 is statistically significant at $\alpha = 0.05$. From statistical point of view, $R_E^2 = 0.015$ implies that the regression model has explained about 1.5% of the total variation in the dependent variables after adjusting for number of explanatory variables and sample size in the model.

The new decision rule for F-test based on Ezekiel's adjusted R-squared presented in this study was developed based on the classical decision rule for F-test. The analysis above provides an insight into the corresponding adjusted R-squared value for a statistically significant regression model based on the significance level in the classical decision rule for F-test. From the numerical example shown above, a statistically significant regression model based on classical decision rule for F-test having an adjusted R-squared value slightly above 1.5%. However, the small value in $(R_E^2)_{CV}$ shows that there is an uncertainty as to whether an adjusted R-squared value that is slightly above 0.015 is sufficient to represent a statistically significant regression model.

5. Conclusion

In this study, we introduced a new decision rule for the F-test in regression analysis based on Ezekiel's adjusted R-squared. The influence of various parameters (alpha value, number of explanatory variables and sample size) on the critical value $(R_E^2)_{CV}$ in the new decision rule were also evaluated.

The results indicated that the values of $(R_E^2)_{CV}$ based on $\alpha = 0.01$ are higher than those based on $\alpha = 0.05$. This is due to the rejection region in F-test based on $\alpha = 0.01$ is smaller than the rejection region based on $\alpha = 0.05$.

The results also showed that the values of critical value tend to increase when the model complexity increases. The model complexity increases when the number of explanatory variables in the linear regression model increases.

However, greater attention should be given to the influence of sample size on the critical value in the new decision rule. From this study, the values of the critical value $(R_E^2)_{CV}$ tend to decrease when sample size increases. Since the new decision rule for the F-test based on Ezekiel's adjusted R-squared, as proposed in this study, was derived from the classical F-test decision rule, the observation of a low value in $(R_E^2)_{CV}$ raises uncertainty regarding whether such a value, typically regarded as small, is sufficient to indicate a statistically significant regression model. While there are no definitive rules governing the interpretation of adjusted R-squared values, it is important to recognize that their interpretation is inherently dependent on the specific context of the study. For instance, an adjusted R-squared value of 0.5 may be regarded as relatively high within the social sciences where the studies often involve human behavior. In contrast, the same value might be deemed insufficient in the context of controlled engineering experiments in which greater explanatory power is typically anticipated (Milton & Arnold 2004). Consequently, the assessment of what constitutes an acceptable or meaningful adjusted R-squared value should be guided by the judgment of domain experts who are familiar with the conventions and expectations within their respective fields.

The results of this study correspond with prior research in which the conclusions derived from hypothesis testing are regarded as uncertain and have been subject to criticism (Nuzzo 2014; Woolston 2015). The issue is rooted in the null hypothesis significance testing paradigm and discussions surrounding this scientific methodology continue to take place. As an alternative, some studies have called for the abandonment of statistical significance and the dropping of the null hypothesis significance testing paradigm (Maneejuk & Yamaka 2021). However, this issue has also been addressed and examined in other academic investigations (Spreckelsen 2018; McShane *et al.* 2019; Imbens 2021; Hussey 2023). In conclusion, the null hypothesis significance testing procedure, including its application in linear regression analysis, should not be dismissed outright. Rather, it should be regarded as one of several screening tools to be interpreted in conjunction with other relevant factors including prior evidence, data quality, design of study, reproducibility, analytical methods, plausibility of the underlying mechanism, reporting practices, novelty of the findings, real-world benefits and costs, as well as other domain-specific considerations. For future research direction, the issue of null hypothesis significance testing paradigm could also be extended to goodness-of-fit test or other types of models such as logistic regression model.

Acknowledgments

The author is grateful to Universiti Teknologi MARA for the support in research and publication.

References

Arpaia P., Carleo G., Gargiulo L., Mancino F., Moccaldi N. & Taglialatela M. 2022. Insulin absorption assessment based on bioimpedance measurements: Model optimization for improving accuracy in clinical application. *2022 International Workshop on Impedance Spectroscopy (IWIS)*, pp. 56-60.

Ayabe C.R. 1985. Multicrossvalidation and the jackknife in the estimation of shrinkage of the multiple coefficient of correlation. *Educational and Psychological Measurement* **45**(3): 445-451.

Ebiwonjumi A., Chifurira R. & Chinhamu K. 2023. A robust principal component analysis for estimating economic growth in Nigeria in the presence of multicollinearity and outlier. *Journal of Statistics Applications & Probability* **12**(2): 611-627.

Hittner J.B. 2020. Ezekiel's classic estimator of the population squared multiple correlation coefficient: Monte Carlo-based extensions and refinements. *The Journal of General Psychology* **147**(3): 213-227.

Huberty C.J. & Mourad S.A. 1980. Estimation in multiple correlation/prediction. *Educational and Psychological Measurement* **40**(1): 101-112.

Hsu C.W., Chan M.J., Weng C.H., Tsai T.Y., Yen T.H. & Huang W.H. 2025. Environmental PM2.5 exposure: An ignored factor associated with blood cadmium level in hemodialysis patients. *Therapeutics and Clinical Risk Management* **21**: 1-13.

Hussey, I. 2023. A systematic review of null hypothesis significance testing, sample sizes, and statistical power in research using the Implicit Relational Assessment Procedure. *Journal of Contextual Behavioral Science* **29**: 86-97.

Imbens G.W. 2021. Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives* **35**(3): 157-174.

Karch J. 2020. Improving on adjusted R-squared. *Collabra: Psychology* **6**(1): 1-11.

Kennedy E. 1988. Estimation of the squared cross-validity coefficient in the context of best subset regression. *Applied Psychological Measurement* **12**(3): 231-237.

Kromrey J.D. & Hines C.V. 1995. Use of empirical estimates of shrinkage in multiple regression: A caution. *Educational and Psychological Measurement* **55**(6): 901-925.

Krus D.J. & Fuller E.A. 1982. Computer assisted multicrossvalidation in regression analysis. *Educational and Psychological Measurement* **42**(1): 187-193.

Maneejuk P. & Yamaka W. 2021. Significance test for linear regression: how to test without P-values?. *Journal of Applied Statistics* **48**(5): 827-845.

McClave J.T., Benson P.G. & Sincich T. 2001. *Statistics for Business and Economics*. New Jersey: Prentice Hall.

McShane B.B., Gal D., Gelman A., Robert C. & Tackett J.L. 2019. Abandon statistical significance. *The American Statistician* **73**(1): 235-245.

Meng W., He C., Yan Q., Guo D., Li R., Tang J., Liao Y. & Chen Z. 2021. Application of p-Value Approach in Inversion of Initial Geo-Stress Field of Rock Masses. *China Railway Science* **42**(1): 71–79

Milton J.S. & Arnold J.C. 2004. *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*. New York: McGraw-Hill.

Mozumder C., Reddy K.V. & Pratap D. 2013. Air pollution modeling from remotely sensed data using regression techniques. *Journal of the Indian Society of Remote Sensing* **41**: 269-277.

Ng S.F. 2025. A theoretical and simulation study in formulating the optimal biasing parameter for an almost unbiased regression estimator. *MATEMATIKA* **41**(1): 57-76.

Nuzzo R. 2014. Scientific method: statistical errors. *Nature* **506**: 150-152.

Raju N.S., Bilgic R., Edwards J.E. & Fleer P.F. 1999. Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement* **23**(2): 99-115.

Spreckelsen T.F. 2018. Changes in the field: banning p-values (or not), transparency, and the opportunities of a renewed discussion on rigorous (quantitative) research. *Child and Adolescent Mental Health* **23**(2): 61-62.

Walpole R.E., Myers R.H., Myers S.L. & Ye K. 2007. *Probability and Statistics for Engineers and Scientists*. New Jersey: Pearson Prentice Hall.

Wang S. & Cui H. 2013. Generalized F test for high dimensional linear regression coefficients. *Journal of Multivariate Analysis* **117**: 134-149.

Wherry R.J. 1931. A new formula for predicting the shrinkage of the coefficient of multiple correlation. *The Annals of Mathematical Statistics* **2**(4): 440-457.

Woolston C. 2015. Psychology journal bans P values. *Nature* **519**: 9.

Yin P. & Fan X. 2001. Estimating R2 shrinkage in multiple regression: A comparison of different analytical methods. *The Journal of Experimental Education* **69**(2): 203-224.

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Johor Branch, Pasir Gudang Campus
81750 Masai
Johor, MALAYSIA
*E-mail: ngsetfoong061@uitm.edu.my**

Received: 30 October 2025
Accepted: 7 January 2026

*Corresponding author