



RESEARCH ARTICLE

COMPARISON OF SOME EFFECT SIZE MEASURES IN SIMPLE AND MULTIPLE
LINEAR REGRESSION MODELS

Soner YİĞİT ^{1,*} 

¹ Biometry and Genetic Unit, Animal Science Department, Faculty of Agriculture, Canakkale Onsekiz Mart University, Canakkale, Turkey

ABSTRACT

It is very important that the results of the statistical analysis are understandable. Therefore, while reporting the results of statistical analysis, some effect size measures should be given along with the P-value. In this study, some effect size measure (R^2 (Eta-Squared), R_ϵ^2 (Epsilon-Squared) and R_ω^2 (Omega-Squared)) were compared in terms of their performance (bias) in simple and multiple linear regression models. For this purpose, Monte Carlo simulation technique was used. Results of simulation runs showed that very small sample sizes caused the predictions of all three effect size measures to deviate significantly. On the other hand, deviations of all three effect size measures gradually decreased due to increased sample sizes. However, the R_ϵ^2 and R_ω^2 estimates were quite unbiased when compared to R^2 regardless of experimental conditions. Thus, it could be concluded that reporting R_ϵ^2 or R_ω^2 is more appropriate to evaluate the practical significance along with P-values in simple and multiple linear regression models.

Keywords: Practical significance, Eta-squared, Epsilon-squared, Omega-squared, Regression analysis, Simulation

1. INTRODUCTION

Linear regression analysis is a quite commonly used statistical method for modeling the relationship between a continuous dependent variable and a simple or multiple continuous independent variable(s) [1-3]. However, it does not show how much of the variation that occurred in the dependent variable is caused by the independent variable or variables. In other words, as in all other methods that test statistical significance, while testing the effects of independent variables on the dependent variable statistically, it does not describe the practical significance of these effects [4-7].

Hays [8] states that practical significance (effect size) is as significant as the test of hypothesis. Today, a good portion of scientific journals encourages reporting some effect size measures alongside the P-value which shows statistical significance because reporting only statistical significance may not provide the information researchers need. To make the results more conceivable, reporting some effect size measures alongside statistical significance would be useful for readers [9].

There are several effect size measures developed for this purpose [8, 10-14]. However, deciding which effect size measures are to be used is critical because to what level the effect size measures represent the population, which means how unbiased they are, has not been analyzed in detail.

Advancing from this point of view, this study aims to compare R^2 , R_ϵ^2 , and R_ω^2 , which are the most popular effect size measures offered for linear models, in terms of their performances under different experimental conditions. Thus, the most suitable and reliable effect size measure that could be used in regression analysis models could be determined.

2. MATERIAL AND METHOD

The materials of this study consist of random numbers generated from Multivariate Normal Distribution by Monte Carlo simulation technique. “mvrnorm” function of the R-Project (2018) program is used to generate the numbers. In this study, R^2 , R_ε^2 ve R_ω^2 are compared in terms of their performances (bias) under different experimental conditions such as correlation structures, independent variable, and sample size. Experimental conditions considered in the study are given in Table 1.

Table 1. Considered Experimental Design for Simple and Multiple Regression Analysis

Statistical Model	$Y=\alpha+\beta X+\varepsilon$	$Y=\alpha+\beta_1 X_1+\beta_2 X_2+\dots+\beta_p X_p+\varepsilon$
The correlation coefficient among independent variables ($\rho_{X_i X_j}$)	-	0.1, 0.3 and 0.5
The correlation coefficient between dependent and independent variables ($\rho_{Y X_i}$)	0.1, 0.3 and 0.5	0.1, 0.3 and 0.5
number of independent variable (p)	1	1,2,...,10
n/(p+1)	2,3,...,10 and 20	2,3,...,10 and 20
number of simulation (n_{sim})	1000000	1000000

Deviations of sample effect size measures from population effect size measures are considered as performance criteria. To calculate the performances of effect size measures of each experimental condition considered in this study, the following steps were applied.

- 1- Determined size (n), 1000000 (n_{sim}) samples are taken from population.
- 2- R_i^2 , $R_{\varepsilon_i}^2$ ve $R_{\omega_i}^2$ is calculated for each taken sample.
- 3- Then, $R^2 = (\sum_{i=1}^{n_{sim}} R_i^2)/n_{sim}$, $R_\varepsilon^2 = (\sum_{i=1}^{n_{sim}} R_{\varepsilon_i}^2)/n_{sim}$ ve $R_\omega^2 = (\sum_{i=1}^{n_{sim}} R_{\omega_i}^2)/n_{sim}$ is calculated.
- 4- Finally, performances of three effect size measures are determined by $bias(R^2) = (R^2 - \rho^2)$, $bias(R_\varepsilon^2) = (R_\varepsilon^2 - \rho^2)$ and $bias(R_\omega^2) = (R_\omega^2 - \rho^2)$.

2.1. Effect Size Measures

To predict the population effect size measure (ρ^2), several effect size measures were developed. In this study, very commonly known effect size measures such as R^2 , R_ε^2 ve R_ω^2 are considered.

$$R^2 = \frac{SS_{Regression}}{SS_{Total}} \quad (1), [1, 3]$$

$$R_\varepsilon^2 = \frac{SS_{Regression} - p \cdot MS_{Error}}{SS_{Total}}, \text{ if } R_\varepsilon^2 < 0, R_\varepsilon^2 = 0 \quad (2), [10]$$

$$R_\omega^2 = \frac{SS_{Regression} - p \cdot MS_{Error}}{SS_{Total} + MS_{Error}}, \text{ if } R_\omega^2 < 0, R_\omega^2 = 0 \quad (3), [8]$$

Where, SS_{Total} : Total sum of squares, $SS_{Regression}$: Sum of squares of regression, MS_{Error} : Error mean square and p: Degrees of freedom of regression

2.2. Determining Population Effect Size

For each experimental condition considered in this study, the effect size of the population (ρ^2) was determined. Since there are one continuous independent and one continuous dependent variable in Simple Linear Regression Analysis, the effect size of the population can be calculated as $\rho^2 = \rho_{YX}^2$. Population effect size measures (ρ^2) calculated for simple linear regression analysis are given in Table 2.

Table 2. Population effect size (%) for Simple Regression Model

	$\rho_{YX} = 0.1$	$\rho_{YX} = 0.3$	$\rho_{YX} = 0.5$
ρ^2	1.00	9.00	25.00

For Multiple Linear Regression Analysis, population effect sizes (ρ^2) can be found easily through the below equations.

$$\rho^2 = \sum_{i=1}^p \beta_i \rho_{YX_i} \tag{4}$$

$$\begin{aligned} \beta_1 &= r^{11} \rho_{YX_1} + r^{12} \rho_{YX_2} + \dots + r^{1p} \rho_{YX_p} \\ \beta_2 &= r^{21} \rho_{YX_1} + r^{22} \rho_{YX_2} + \dots + r^{2p} \rho_{YX_p} \\ &\vdots \\ \beta_p &= r^{p1} \rho_{YX_1} + r^{p2} \rho_{YX_2} + \dots + r^{pp} \rho_{YX_p} \end{aligned} \tag{5}$$

Where r^{ij} : element of the inverse of $\rho_{X_i X_j}$ [2].

Population effect sizes (ρ^2) calculated this way are given in Table 3.

Table 3. Population effect size (%) for Multiple Regression Model

$\rho_{X_i X_j}$	ρ_{YX_i}	p=2	p=3	p=4	p=5	p=6	p=7	p=8	p=9	p=10
0.1	0.1	1.82	2.50	3.08	3.57	1.00	4.38	4.71	5.00	5.26
	0.3	16.36	22.50	27.69	32.14	36.00	39.37	42.35	45.00	47.37
	0.5	45.45	62.50	76.92	89.29	100.00	100.00	100.00	100.00	100.00
0.3	0.1	15.38	1.88	2.11	2.27	2.40	2.50	2.58	2.65	2.70
	0.3	13.85	16.88	18.95	20.45	21.60	22.50	23.23	23.82	24.32
	0.5	38.46	46.88	52.63	56.82	60.00	62.50	64.52	66.18	67.57
0.5	0.1	1.33	1.50	1.60	1.67	1.71	1.75	1.78	1.80	1.82
	0.3	12.00	13.50	14.40	15.00	15.43	15.75	16.00	16.20	16.36
	0.5	33.33	37.50	40.00	41.67	42.86	43.75	44.44	45.00	45.45

3. RESULTS

In this study, three different correlation structures and ten different sample sizes in Simple Linear Regression Model; nine different correlation structures, nine different independent variables, and ten different sample sizes in Multiple Linear Regression Model; in total 900 different experimental conditions were considered. Results for Simple Linear Regression Model are given in figure 1, and for Multiple Linear Regression Model are given in figure 2-4 in detail.

3.1. Results for Simple Linear Regression Model

In all experimental conditions considered when $\rho_{YX_i}=0.1$, deviations of predictions of three effect size measures decreased, as expected, in correlation with the increase in sample size. Generally, R^2 gave more deviant predictions compared to R^2_ϵ ve R^2_ω . R^2_ϵ ve R^2_ω created acceptable deviations (less than 5%) when $n/(p + 1) \geq 5$. However, R^2 was below this level when $n/(p + 1) > 10$. On the other hand, when $n/(p + 1) < 5$, deviations that occurred in all three effect size measures were significantly high. This condition became more obvious as the sample size decreased. For example, when $n/(p + 1) = 2$, deviations were R^2 32.64%, R^2_ϵ 18.52% and R^2_ω 16.51% (Figure 1).

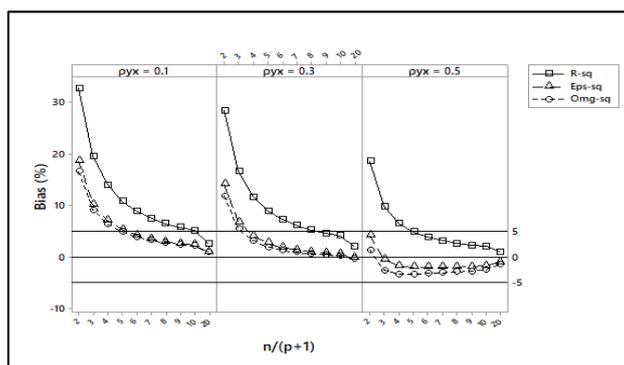


Figure 1. Bias for Simple Linear Regression Models

Depending on the increase in ρ_{YX_i} , deviations in the predictions of all three effects size measures gradually decreased. For example, when $n/(p + 1) = 2$, deviations of the predictions of R^2 , R_ξ^2 and R_ω^2 were respectively 32.64%, 18.52% and 16.51% for $\rho_{YX_i}=0.1$; 28.32%, 14.02% and 11.80 for $\rho_{YX_i}=0.3$; and 18.57%, 4.14% and 1.39% for $\rho_{YX_i}=0.5$.

3.2. Results for Multiple Linear Regression Model

3.2.1. Results for $\rho_{X_iX_j}=0.1$

When there were weak relations ($\rho_{YX_i}=0.1$) between independent variables and dependent variable, regardless of the experimental conditions, R_ξ^2 and R_ω^2 made quite unbiased predictions compared to R^2 . This condition became more significant as the number of variables increased (Figure 2). On the other hand, as the sample size increased, deviations in the predictions of R^2 increased more. Additionally, when $n/(p + 1) < 20$, deviations of the predictions made by R^2 never went below 5.00% under any experimental condition.

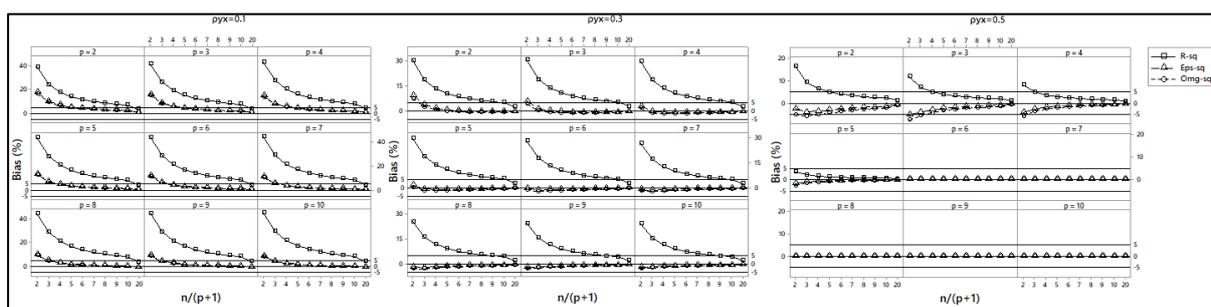


Figure 2. Bias for Multiple Linear Regression Models When $\rho_{X_iX_j}=0.1$ and $\rho_{YX_i}=0.1, 0.3$ and 0.5

When $\rho_{YX_i}=0.3$, deviations of the predictions of all three effect size measures were seen to gradually decrease. However, as long as $n/(p + 1) \leq 9$, deviations of predictions of R^2 did not go below 5.00%. When $\rho_{YX_i}=0.3$, R_ξ^2 and R_ω^2 made unbiased predictions under almost all experimental conditions (except $p=2$ and $n/(p + 1) = 2$) (Figure 2). As the relation between independent variables and dependent variable got stronger ($\rho_{YX_i}=0.5$), it was seen that predictions decreased significantly (Figure 2). When $p \geq 6$, all effect size measures gave almost unbiased predictions (bias $\cong 0$).

3.2.2. Results for $\rho_{X_i X_j}=0.3$

When $\rho_{Y X_i}=0.1$, as long as the sample sizes were not extremely large ($n/(p + 1) < 20$), R^2 gave quite deviant predictions (Figure 3). Except for $n/(p + 1) \leq 3$, R_ϵ^2 and R_ω^2 generally made quite unbiased predictions. This condition became more significant as the number of independent variables (p) increased. However, R^2 was affected negatively by the increase in the number of independent variables. Although deviations in the predictions decreased to some degree when there were moderate level relations between independent variables and dependent variable ($\rho_{Y X_i}=0.3$), R^2 never went below 5.00% deviation under any experimental conditions (except for $n/(p + 1) < 20$) (Figure 3) R_ϵ^2 and R_ω^2 made quite unbiased predictions as long as $n/(p + 1) > 2$.

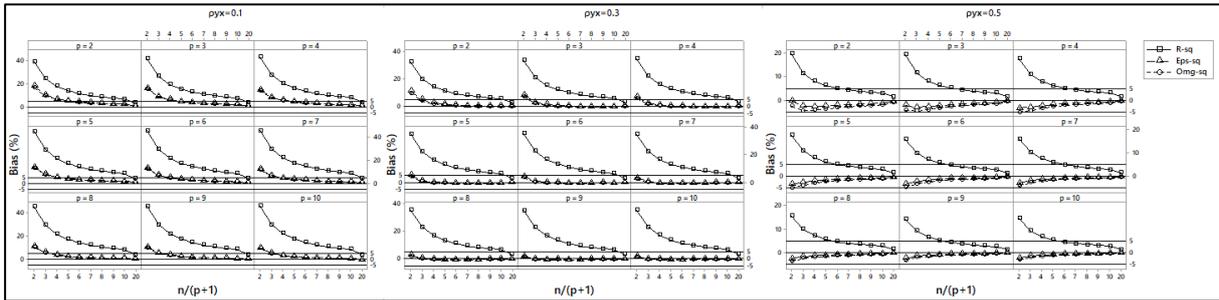


Figure 3. Bias for Multiple Linear Regression Models When $\rho_{X_i X_j}=0.3$ and $\rho_{Y X_i}=0.1, 0.3$ and 0.5

For $\rho_{Y X_i}=0.5$, regardless of the sample size, deviations of the predictions of R_ϵ^2 and R_ω^2 were always below 5.00%. Although R^2 was also affected positively from the increase of $\rho_{Y X_i}$, deviations in the predictions went below 5.00% only after $n/(p + 1) > 5$ (Figure 3).

3.2.3. Results for $\rho_{X_i X_j}=0.5$

When the relation between independent variables and dependent variable weak ($\rho_{Y X_i}=0.1$), relation between independent variables ($\rho_{X_i X_j}$) did not affect the accuracy of the predictions (Figure 4). Although the deviations of the predictions of R_ϵ^2 and R_ω^2 were over 5.00% when the sample sizes were small ($n/(p + 1) \leq 4$), this decreased as the number of variables increased. However, deviations of the predictions of R_ϵ^2 and R_ω^2 did not go below 5.00% as long as $n/(p + 1) \leq 3$. R^2 made predictions with deviations over 5.00% under all experimental conditions when $n/(p + 1) < 20$. This became more obvious as the sample size decreased.

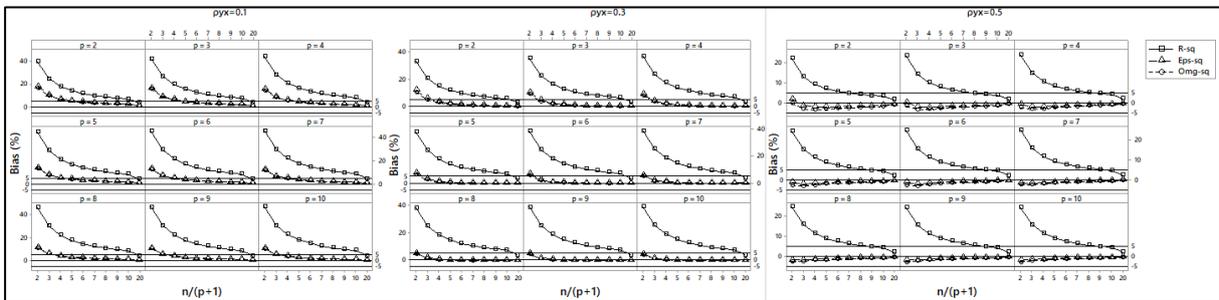


Figure 4. Bias for Multiple Linear Regression Models When $\rho_{X_i X_j}=0.5$ and $\rho_{Y X_i}=0.1, 0.3$ and 0.5

Although the increase of ρ_{YX_i} (0.3 and 0.5) provided the deviations of the predictions of R^2 decrease a little, it was determined that there were deviations over 5.00% almost under all experimental conditions (except $n/(p + 1) = 20$). Notwithstanding, R_ε^2 and R_ω^2 gave quite unbiased results, and this became more apparent as the ρ_{YX_i} increased (Figure 4).

4. DISCUSSION

Linear regression analysis is a quite commonly used method for modeling the relationship between a dependent variable and independent variable(s). However, it does not provide any information about the practical significance (explained variance) of independent variables. Thus, reporting only P-value with the acquired regression equation is not enough. To make the results more understandable, some effect size measures should be reported along with the P-value. There are several effect size measures developed for this purpose. Besides, R^2 is used as model selection criteria as well. Of the considered models, the one with the higher R^2 is considered better. However, whether R^2 is reliable under the considered experimental conditions is quite important. It is an important drawback that the performances of effect size measures, which could be used both for practical significance and as model selection criteria, are not known. Revealing the performances of effect size measures under several experimental conditions will guide researchers about which effect size measure to be reported. Advancing from this point of view, in this study, R^2 , R_ε^2 and R_ω^2 were compared in terms of their performances under a wide range of experimental conditions.

Generally, in the literature, performances of effect size measures are presented in ANOVA models. Although Keselman [16] states that R^2 , R_ε^2 and R_ω^2 gives similar results in One-Way ANOVA, many studies show that R_ε^2 and R_ω^2 are more unbiased than R^2 [4, 17-20]. This study reveals that R_ε^2 and R_ω^2 are generally significantly unbiased compared to R^2 in Regression Models. However, very small sample sizes ($n/(p + 1) = 2$) caused the predictions of all three effect size measures to deviate significantly. On the other hand, deviations of all three effect size measures gradually decreased due to increased sample sizes.

In practice, as independent variables are added to the model, there is a belief that R^2 increases artificially [21, 22]. It is natural that R^2 increases as independent variables are added to the model because explained variation should increase unless the relationship between independent variables and dependent variable is zero. The important thing at this point is to what degree R^2 , which is calculated from the sample, represents ρ^2 , which is population effect size. Results of this study reveal that the number of independent variables added to the model is the least affecting factor of the deviations in predictions. Furthermore, when the relationship between independent variables and dependent variable increases, R^2 is positively affected by the increase of the number of independent variables. An increase in the number of the independent variable always affects R_ε^2 and R_ω^2 positively.

While Keselman [16] reports that the bias of the effect size measures increases as the population effect size increases in One-Way ANOVA, Yiğit and Mendes [20] report that bias of the effect size measures decreases as the population effect size increases. In this study, similar results with Yiğit and Mendes [20] were obtained. Thus, in reality, if there are strong relations between independent variables and dependent variable, deviations of the predictions of the effect size measures decrease.

Low-level relations between independent variables affect all three effect size measures negatively. However, R_ε^2 and R_ω^2 give quite unbiased results compared to R^2 .

Under all experimental conditions considered in this study, R_ε^2 and R_ω^2 provided significantly reliable results compared to R^2 .

5. CONCLUSION

Statistical results obtained from a study are very important. However, analyzing only statistical significance may result in misunderstandings. For instance, there is a belief in practice that results are more important as the P-value decreases. Nevertheless, statistical significance is associated with the sample size. As the sample size increases, statistical significance will increase as well. In this case, interpretations made considering only P-value may cause misunderstandings. While reporting the results of a study, reporting effect size measures that represent practical significance along with the P-value will be more understandable. The important thing here is knowing which effect size measure represents the population better.

In this study, the most common effect size measures are compared in Simple and Multiple Linear Regression Models under several experimental conditions. Results of the simulations show that R^2 , which is reported commonly, is not reliable. On the other hand, R^2_{ξ} and R^2_{ω} , although they give deviant results under some experimental conditions, give generally quite unbiased results.

Another important point to highlight is that predictions are quite deviant for all three effect size measures when there is a weak relation between independent variables and dependent variable. Researchers should take this into consideration.

As conclusion, reporting effect size measures like R^2_{ξ} and R^2_{ω} along with P-value would provide the research results to be more understandable.

CONFLICT OF INTEREST

The author stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

- [1] Zar JH. Biostatistical Analysis. 5th ed. New Jersey, USA: Pearson, 2014.
- [2] Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for The Behavioral Sciences. 3rd ed. New Jersey, USA: Lawrence Erlbaum, 2003.
- [3] Draper NR, Smith H. Applied Regression Analysis. 3rd ed. New York, USA: John Wiley & Sons, 2014.
- [4] Glass GV, Hakstian, AR. Measures of association in comparative experiments: Their development and interpretation. *Am Educ Res J* 1969; 6: 403–414.
- [5] Cohen J. Things I have learned (so far). *Am Psychol* 1990; 45: 1304–1312.
- [6] Klein DF. Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research. *Am J Psychiat* 2005; 162: 643-a-644.
- [7] Sullivan GM, Richard F. Using Effect Size-or why the P value is not enough. *J Grad Med Educ* 2012; 4: 279-282.
- [8] Hays W. Statistics for Psychologists. 1st ed., New York, USA: Rinehart Winston, 1963.
- [9] Thompson B. Editorial policies regarding statistical significance testing: Three suggested reforms. *Educ Researcher* 1996; 25: 26-30.

- [10] Kelley TL. An unbiased correlation ratio measure. *Proc Natl Acad Sci USA* 1935; 21: 554– 559.
- [11] Maxwell SE, Camp CJ, Arvey RD. Measures of strength of association: A comparative examination. *J Appl Psychol* 1981; 6: 525–534.
- [12] Keppel G. *Design and Analysis: A Researcher’s Handbook*. 1st ed. New Jersey, USA: Prentice Hall, 1982.
- [13] Olejnik S, Algina J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemp Educ Psychol* 2000; 25: 241–286.
- [14] Grissom R, Kim J. *Effect Sizes for Research: A Broad practical approach*. New Jersey, USA: Lawrence Erlbaum, 2005.
- [15] Core R, Team R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2018.
- [16] Keselman HJ. A Monte carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Can Psychol Rev* 1975; 16: 44–48.
- [17] Carroll RM, Nordholm LA. Sampling characteristics of Kelley’s epsilon and Hays’ omega. *Educ Psychol Meas* 1975; 35: 541–554.
- [18] Okada K. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika* 2013; 40: 129–147.
- [19] Skidmore ST, Thompson B. Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behav Res Methods* 2012; 45: 536–546.
- [20] Yigit S, Mendes M. Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study. *Revstat-Stat J* 2018; 16: 295-313.
- [21] Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010; 107: 776-882.
- [22] Doulah MSU. Robust coefficients of determination: A measure of goodness of fit. *Int J Eng Res* 2013; 4: 1715-1717.