

STATISTICAL ASSUMPTION TESTING IN MULTIPLE LINEAR REGRESSION FOR EDUCATIONAL RESEARCH: EVIDENCE FROM A SECONDARY SCHOOL SURVEY IN INDONESIA

^{*1}Fitriyanti, ²Meli Puspita, ³Fajri Ismail, ⁴Muhammad Win Afghani

^{*1,2,3,4}Universitas Islam Negeri Raden Fatah Palembang

Email: ^{*1}fitriyantididik@gmail.com, ²melipuspita45@guru.sd.belajar.id,

³fajriismail_uin@radenfatah.ac.id, ⁴muhammadwinafgani_uin@radenfatah.ac.id

Abstract

Multiple linear regression is widely used in educational research to examine the statistical association of several predictors with an outcome. Its interpretation, however, depends on appropriate model specification and diagnostic evaluation. This study examined a two-predictor regression model of students' learning outcomes and evaluated the statistical assumptions preceding model interpretation. A quantitative survey design was conducted at SMA Negeri 1 Rantau Bayur. The sample comprised 40 students selected purposively. Data were collected using Likert-scale questionnaires whose validity and reliability had been assessed and were analyzed using IBM SPSS Statistics. The analysis included descriptive statistics, distributional diagnostics, homoscedasticity, multicollinearity, heteroskedasticity diagnostics, and multiple linear regression. The reported results showed Shapiro–Wilk p values above .05 for X_1 , X_2 , and Y ; a Levene statistic of 1.244 ($p = .272$); tolerance of .712 and VIF of 1.404 for both predictors; and no apparent heteroskedastic pattern in the residual scatterplot. The regression model yielded positive coefficients for X_1 ($B = 0.413$, $SE = 0.098$, $p < .001$) and X_2 ($B = 0.287$, $SE = 0.081$, $p = .001$), with $R = .742$, $R^2 = .551$, and adjusted $R^2 = .526$. Thus, the two predictors jointly accounted for 55.1% of the observed variance in learning outcomes in this sample. The study highlights the importance of regression diagnostics while emphasizing that purposive sampling and a single-school sample constrain generalizability and that the observed associations should not be interpreted causally.

Keywords: multiple linear regression, regression diagnostics, statistical assumptions, learning outcomes, quantitative educational research, purposive sampling, Indonesia

Abstrak

Regresi linier berganda banyak digunakan dalam penelitian pendidikan untuk mengkaji hubungan statistik antara beberapa variabel prediktor dengan variabel hasil. Namun, interpretasinya bergantung pada spesifikasi model yang tepat dan evaluasi diagnostik. Penelitian ini mengkaji model regresi dengan dua variabel prediktor terhadap hasil belajar siswa serta mengevaluasi asumsi-asumsi statistik yang mendasari interpretasi model. Desain survei kuantitatif dilakukan di SMA Negeri 1 Rantau Bayur. Sampel terdiri dari 40 siswa yang dipilih secara purposif. Data dikumpulkan menggunakan kuesioner skala Likert yang validitas dan reliabilitasnya telah dievaluasi, kemudian dianalisis menggunakan IBM SPSS Statistics. Analisis tersebut mencakup statistik deskriptif, diagnostik distribusi, homoskedastisitas, multikolinearitas, diagnostik

*heteroskedastisitas, dan regresi linier berganda. Hasil yang dilaporkan menunjukkan nilai p Shapiro–Wilk di atas 0,05 untuk X_1 , X_2 , dan Y ; statistik Levene sebesar 1,244 ($p = 0,272$); toleransi sebesar 0,712 dan VIF sebesar 1,404 untuk kedua prediktor; serta tidak terdapat pola heteroskedastisitas yang jelas pada diagram sebar residu. Model regresi menghasilkan koefisien positif untuk X_1 ($B = 0,413$, $SE = 0,098$, $p < 0,001$) dan X_2 ($B = 0,287$, $SE = 0,081$, $p = 0,001$), dengan $R = 0,742$, $R^2 = 0,551$, dan R^2 yang disesuaikan = 0,526. Dengan demikian, kedua prediktor tersebut secara bersama-sama menjelaskan 55,1% varians yang diamati pada hasil belajar dalam sampel ini. Penelitian ini menyoroti pentingnya diagnostik regresi sekaligus menekankan bahwa pengambilan sampel secara purposif dan sampel yang hanya berasal dari satu sekolah membatasi generalisasi, serta bahwa hubungan yang diamati tidak boleh ditafsirkan secara kausal. **Kata kunci:** regresi linier berganda, diagnostik regresi, asumsi statistik, hasil belajar, penelitian pendidikan kuantitatif, pengambilan sampel secara purposif, Indonesia*

INTRODUCTION

Multiple linear regression is a central analytical tool for examining how several predictors are statistically related to an educational outcome. It allows researchers to estimate partial regression coefficients while accounting for the other predictors in the model (Cohen et al., 2003; Fox, 2016; Harrell, 2015). Its value in educational research therefore depends not only on obtaining coefficients and significance tests but also on demonstrating that the model is appropriately specified and that its inferential assumptions have been considered.

Regression is a model of the conditional mean of an outcome and its error structure. Relevant considerations include linearity, independence, homoscedasticity, absence of problematic multicollinearity, and—when conventional small-sample inference is used—an appropriate distribution of model errors (Casson & Farmer, 2014; Montgomery et al., 2021; Wooldridge, 2016). Residual analysis is consequently important because it can reveal nonlinearity, heteroskedasticity, influential observations, and departures from model assumptions (Nobre & Singer, 2007).

Normality is frequently handled imprecisely in applied research. For linear regression, the relevant distributional concern is principally the error or residual distribution for finite-sample inference, rather than a requirement that each observed variable itself be normally distributed (Schmidt & Finan, 2018). Recent methodological evidence has documented the common practice of testing raw variables instead of regression residuals, highlighting the need for more precise reporting (Sato et al., 2025). The Shapiro–Wilk test is useful as a formal diagnostic, but it should be interpreted alongside graphical evidence and the model context (Shapiro & Wilk, 1965; Arnastauskaitė et al., 2021).

Homoscedasticity is another important assumption because systematic changes in residual variance can affect standard errors and inferential validity (Yang et al., 2019). Multicollinearity creates a separate problem: high correlation among predictors can inflate coefficient uncertainty and complicate interpretation of individual effects (Kim,

2019; Beckstead, 2012). These issues indicate that assumption testing should be part of the analytical reasoning rather than a mechanical checklist.

Transparent reporting is also increasingly emphasized in quantitative research. The APA Journal Article Reporting Standards recommend clear reporting of sampling, measures, data exclusions, analytic procedures, statistical assumptions, effect estimates, and limitations (Appelbaum et al., 2018). Observational-study reporting guidance similarly stresses transparent description of participants, variables, statistical methods, and limits on generalizability (von Elm et al., 2007). These principles are especially relevant for a study based on 40 purposively selected students.

The present study examines a two-predictor regression model of learning outcomes among students at SMA Negeri 1 Rantau Bayur. The source manuscript reports descriptive statistics, Shapiro–Wilk tests, Levene’s test, tolerance and VIF statistics, a residual scatterplot, regression coefficients, and R^2 . However, the substantive names of X1 and X2 are not specified in the available manuscript. This revision therefore retains the labels X1 and X2 and does not invent construct names.

The methodological gap addressed by this article concerns the interpretation and reporting of regression diagnostics in educational research. The study asks: (1) What are the descriptive characteristics of X1, X2, and Y? (2) What evidence is available concerning distributional characteristics, multicollinearity, and heteroskedasticity? and (3) What associations between X1, X2, and Y are estimated by the regression model? The contribution is positioned as an applied methodological case rather than a causal evaluation of an educational intervention.

Statistical Framework

The model is specified as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where Y denotes learning outcomes, X1 and X2 are the two predictors recorded in the source study, β_0 is the intercept, β_1 and β_2 are partial regression coefficients, and ε represents unexplained variation. A coefficient represents the expected difference in Y associated with a one-unit difference in a predictor while the other predictor is held constant (Cohen et al., 2003; Fox, 2016).

The main diagnostic domains are linearity, independence, homoscedasticity, and multicollinearity. Residual normality is relevant to conventional finite-sample inference, but normality of the raw predictors is not itself a required condition for estimating an ordinary least-squares regression line (Harrell, 2015; Schmidt & Finan, 2018). R^2 represents the proportion of sample outcome variance accounted for by the predictors and should not be interpreted as a causal effect size or as proof of intervention effectiveness (Gelman et al., 2021).

METHOD

Design and Participants

The study used a quantitative survey design and multiple linear regression. Data were collected at SMA Negeri 1 Rantau Bayur. The source manuscript reports 40 students

selected through purposive sampling. The sample was drawn according to criteria related to participation in the relevant learning process. Because purposive sampling was used, the findings describe the observed sample and should not be generalized automatically to all students or schools.

CHARACTERISTIC	REPORTED INFORMATION
SETTING	SMA Negeri 1 Rantau Bayur
SAMPLE	40 students
SAMPLING	Purposive sampling
PREDICTORS	X1 and X2
OUTCOME	Learning outcomes (Y)
SOFTWARE	IBM SPSS Statistics

Measures and Data Collection

Data were collected through Likert-scale questionnaires. The source manuscript states that validity and reliability testing were conducted before use. The available document does not report the number of items, item-level validity evidence, reliability coefficients, or operational definitions of X1 and X2. These details should be inserted from the original instruments and output before submission to a journal requiring full measurement transparency.

Statistical Analysis

The analysis included descriptive statistics, Shapiro–Wilk tests, Levene’s test, multicollinearity diagnostics using tolerance and VIF, visual inspection of a residual scatterplot, and multiple linear regression. In this revision, the Shapiro–Wilk results are treated as evidence concerning the observed variables, not as definitive evidence that the regression residuals are normally distributed. Linearity and independence are recognized as relevant assumptions, but the source document does not provide quantitative diagnostics for them. The regression model was estimated by ordinary least squares, with $\alpha = .05$ used for significance testing.

This reporting strategy follows the principle that statistical analyses should be matched to the actual data and study design and that limitations in diagnostics should be disclosed rather than concealed (Appelbaum et al., 2018).

RESULTS AND DISCUSSION

Descriptive Statistics

VARIABLE	N	MINIMUM	MAXIMUM	MEAN	SD
X1	40	48.80	88.52	69.45	9.32
X2	40	41.32	89.44	64.78	11.85

Y	40	42.30	95.10	71.26	10.74
----------	----	-------	-------	-------	-------

X1 had a mean of 69.45 (SD = 9.32), X2 had a mean of 64.78 (SD = 11.85), and Y had a mean of 71.26 (SD = 10.74). Because the available manuscript does not provide validated cut-off points for the scales, the revised article does not assign unsupported low, moderate, or high categories.

Distributional Diagnostic

Variable	Shapiro–Wilk W	p	Interpretation
X1	0.971	.368	No evidence against normality
X2	0.962	.224	No evidence against normality
Y	0.978	.551	No evidence against normality

All reported p values exceeded .05. Thus, the tests did not provide evidence against normality for the three observed variables. This result should not be equated with proof of normality of regression residuals; residual Q–Q plots or other residual diagnostics would be more directly relevant to the fitted model (Schmidt & Finan, 2018; Sato et al., 2025).

Variance and Multicollinearity Diagnostics

DIAGNOSTIC	STATISTIC	INTERPRETATION
LEVENE’S TEST	1.244, p = .272	No evidence of variance heterogeneity in the reported test
TOLERANCE—X1	.712	No indication of problematic collinearity
VIF—X1	1.404	Low collinearity
TOLERANCE—X2	.712	No indication of problematic collinearity
VIF—X2	1.404	Low collinearity

The tolerance and VIF values indicate little evidence of problematic collinearity between X1 and X2. High multicollinearity can inflate standard errors and make coefficient estimates unstable, whereas the values reported here are comparatively low (Kim, 2019; Beckstead, 2012). Levene’s test is retained as reported in the source, but it is not treated as a substitute for residual-versus-fitted assessment of regression homoscedasticity.

Heteroskedasticity

The source manuscript reports that the residual scatterplot showed randomly dispersed points without an obvious systematic pattern. Accordingly, no clear visual evidence of heteroskedasticity was reported. Because the numerical residual data are unavailable, this conclusion remains qualified. Formal residual-based diagnostics or robust standard errors would strengthen a future analysis (Yang et al., 2019; Wilcox, 2019).

Multiple Linear Regression

Predictor	B	SE	t	p
Constant	8.214	6.521	1.259	.216
X1	0.413	0.098	4.215	< .001
X2	0.287	0.081	3.543	.001
R	R ²	Adjusted R ²	SE of estimate	
.742	.551	.526	7.82	

The reported coefficients yield the equation $\hat{Y} = 8.214 + 0.413X1 + 0.287X2$. Holding X2 constant, a one-unit increase in X1 is associated with an estimated 0.413-unit increase in Y. Holding X1 constant, a one-unit increase in X2 is associated with an estimated 0.287-unit increase in Y. Both coefficients are statistically different from zero at $\alpha = .05$. The R² of .551 indicates that 55.1% of the observed variance in Y is accounted for by X1 and X2 together in this sample.

The unstandardized coefficient for X1 is numerically larger than that for X2, but this does not by itself establish that X1 is more important because the predictors may be measured on different scales. Relative importance would require standardized coefficients or a prespecified relative-importance analysis (Aiken & West, 1991; Grömping, 2007).

The regression estimates represent statistical associations within the sample. They do not demonstrate that increasing either predictor would cause learning outcomes to increase. Causal interpretation would require stronger design and temporal evidence (Gelman et al., 2021; Harrell, 2015).

DISCUSSION

Interpretation of the Regression Model

The main finding is that both predictors were positively associated with learning outcomes and that the model accounted for 55.1% of sample variance. This indicates that the model captures a substantial portion of observed variability, but R² alone does not establish causal validity, correct specification, or external generalizability (Harrell, 2015; Fox, 2016).

Importance of Correct Regression Diagnostics

The methodological contribution is the integration of diagnostics into substantive interpretation. Regression coefficients can be misleading when the model contains influential observations, nonlinearity, heteroskedasticity, or unstable predictor relationships. Residual analysis provides a practical mechanism for checking these issues (Nobre & Singer, 2007; Casson & Farmer, 2014).

The source analysis found $p > .05$ for Shapiro–Wilk tests on X1, X2, and Y. The revised interpretation deliberately avoids stating that these tests prove the regression assumptions. Recent methodological work demonstrates that normality testing should be aligned with the model object of interest, particularly the residuals (Sato et al., 2025).

Multicollinearity and Homoscedasticity

The VIF of 1.404 for both predictors suggests that the two predictors do not exhibit problematic linear redundancy in this sample. This improves the interpretability of their partial coefficients, although VIF thresholds are diagnostic conventions rather than universal laws (Kim, 2019; O’Brien, 2007). The residual scatterplot also did not show an obvious heteroskedastic pattern, but a numerical residual dataset would permit stronger sensitivity analysis (Yang et al., 2019).

Sample Size and Generalizability

The sample of 40 is a substantive limitation. Regression sample-size requirements depend on expected effect size, model complexity, and estimation goals rather than a universal rule (Green, 1991; Austin & Steyerberg, 2015). The purposive sampling strategy further restricts population inference. The study is therefore better understood as a school-level empirical case than as evidence representative of Indonesian secondary students generally.

Methodological Contribution

The article contributes an applied example of transparent regression reporting in an Indonesian educational context. Its value is not the validation of a particular educational intervention, but the demonstration that regression findings should be interpreted jointly with diagnostics, measurement properties, sampling, and design limitations. This positioning is consistent with contemporary quantitative reporting standards (Appelbaum et al., 2018).

Implications

For educational researchers, the findings support more complete reporting of construct definitions, questionnaire items, reliability, missing-data procedures, residual diagnostics, influential observations, confidence intervals, and sample-size justification. For practitioners, the observed associations may identify variables worthy of further investigation, but they should not be translated directly into causal intervention claims.

Future studies should use stronger sampling and longitudinal or experimental designs when causal questions are central.

Limitations and Future Research

The study has six principal limitations: a small single-school sample; purposive rather than probability sampling; incomplete reporting of the substantive identities and measurement properties of X1 and X2; variable-level rather than residual-focused normality testing; limited reported diagnostics for linearity, independence, and influential observations; and a cross-sectional observational design that does not support causal inference. Future research should report complete measurement information, conduct residual-based diagnostics, assess influential cases, justify sample size using effect size and precision, and replicate the model across multiple schools.

CONCLUSION

The study found positive statistical associations of X1 and X2 with learning outcomes in a sample of 40 students at SMA Negeri 1 Rantau Bayur. The reported model produced $B = 0.413$ for X1 and $B = 0.287$ for X2, with $R^2 = .551$ and adjusted $R^2 = .526$. The reported VIF and tolerance values indicated no apparent problematic multicollinearity, while the residual scatterplot showed no obvious heteroskedastic pattern. However, the normality results should be interpreted as evidence concerning the observed variables rather than definitive evidence concerning regression residuals. Given the purposive sample, single-school setting, limited measurement reporting, and observational design, the findings should be treated as sample-specific associations. The principal methodological lesson is that regression results become more credible when diagnostics, measurement, sampling, uncertainty, and limitations are reported transparently.

REFERENCES

- American Psychological Association. (2024). APA style journal article reporting standards (JARS). American Psychological Association. <https://www.apa.org/pubs/journals/resources/apa-style-jars>
- Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Sage.
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Arnastauskaitė, J., Ruzgas, T., & Bražėnas, M. (2021). An exhaustive power comparison of normality tests. *Mathematics*, 9(7), 788. <https://doi.org/10.3390/math9070788>
- Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, 68(6), 627–636. <https://doi.org/10.1016/j.jclinepi.2014.12.014>

- Beckstead, J. W. (2012). Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research*, 47(2), 224–246. <https://doi.org/10.1080/00273171.2012.658331>
- Casson, R. J., & Farmer, L. D. (2014). Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clinical & Experimental Ophthalmology*, 42(6), 590–596. <https://doi.org/10.1111/ceo.12358>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.
- Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Sage.
- Gelman, A., Hill, J., & Vehtari, A. (2021). *Regression and other stories*. Cambridge University Press. <https://doi.org/10.1017/9781139161879>
- Ghozali, I. (2018). *Aplikasi analisis multivariate dengan program IBM SPSS 25* (9th ed.). Badan Penerbit Universitas Diponegoro.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7
- Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2), 139–147. <https://doi.org/10.1198/000313007X188252>
- Harrell, F. E., Jr. (2015). *Regression modeling strategies* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-319-19425-7>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558–569. <https://doi.org/10.4097/kja.19087>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley.
- Nobre, J. S., & Singer, J. M. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, 49(6), 863–875. <https://doi.org/10.1002/bimj.200610341>
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, 673–690. <https://doi.org/10.1007/s11135-006-9018-6>

- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98, 146–151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Midway, S., & White, J. W. (2025). Testing for normality in regression models: Mistakes abound (but may not matter). *Royal Society Open Science*, 12. <https://doi.org/10.1098/rsos.241904>
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Medicine*, 4(10), e296. <https://doi.org/10.1371/journal.pmed.0040296>
- Wilcox, R. R. (2019). Robust regression: Testing global hypotheses about the slopes when there is multicollinearity or heteroscedasticity. *British Journal of Mathematical and Statistical Psychology*, 72(2), 355–369. <https://doi.org/10.1111/bmsp.12152>
- Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.
- Yang, K., Tu, J., & Chen, T. (2019). Homoscedasticity: An overlooked critical assumption for linear regression. *General Psychiatry*, 32(5), e100148. <https://doi.org/10.1136/gpsych-2019-100148>