Paper Review:

"Multiple regression modelling for Mathematics performance: Best model selections"

Elizabeth Bader

Reference

Amir, S. S., Mohd Hashim, S. R., Ag Isha, D. S., & Abdullah, M. K. (2024). Multiple regression modelling for Mathematics performance: Best model selections. *International Journal of Academic Research in Progressive Education and Development*, *13*(3). https://doi.org/10.6007/ijarped/v13-i3/21712

Paper Information (Breakdown)

Title: "Multiple Regression Modelling for Mathematics Performance: Best Model Selections"

Authors: Sitti Sham Amir, Siti Rahayu Mohd Hashim, Dg Siti Nurisya Sahirah Ag Isha, Mohd Khairuddin, Jerry Abdullah

Authors' Associations: various departments of the University of Malaysia, Sabah

Received: 22 Apr, 2024 Revised: 26 May, 2024 Published Online: 28 Jun, 2024

International Journal of Academic Research in Progressive Education & Development, Vol 13, Issue 3, (2024)

http://dx.doi.org/10.6007/IJARPED/v13-i3/21712

Content Relevance

- Multiple regression
- R^2, adjusted R^2
- assumptions of linearity/homogeneity of variance/normality of data
- checking for multicollinearity, outliers

Overview of Research Question

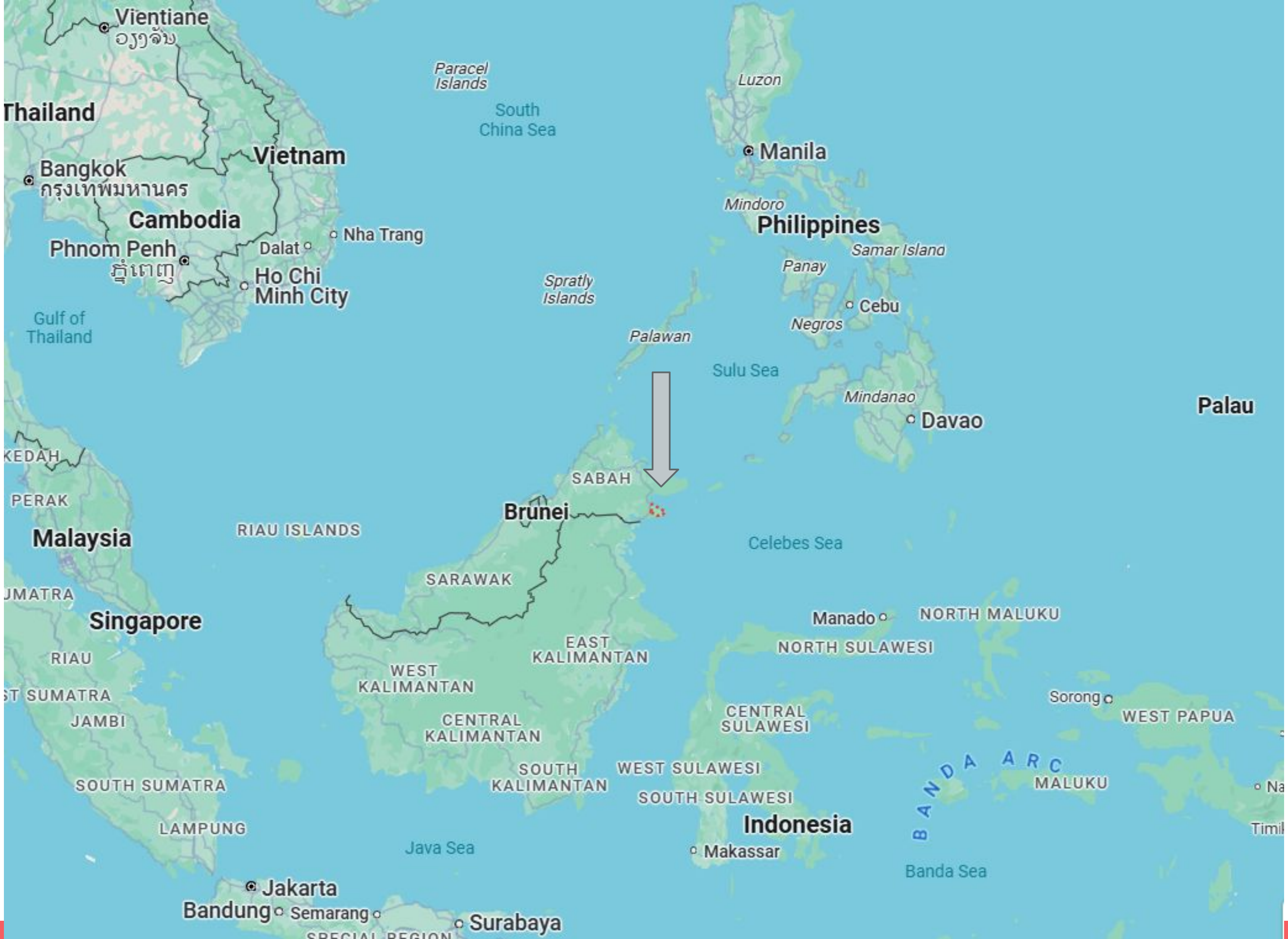
- Researchers compared two explanatory multiple regression models, both of which aimed at identifying factors which improve mathematics achievement among fifth-year students (11 year old children) based on five variables.
- The first model used all variables without utilizing subgroups and did not ensure multiple regression assumptions were met.
- The second model used the same variables but used subsets ("domains" explained later in these slides) and ensured multiple regression assumptions were met. Unsurprisingly, the conclusion was that the second model was more accurate.
- Researchers ultimately identified three key variables which influence math achievement among fifth-year students (to be explained later).
- I additionally sought to understand if the conclusions reached in Malaysia would be similar to conclusions reached in the United States.

Dataset

Sample : 267 fifth-year students in Semporna, Sabah, Malaysia (on the island of Borneo, shared with Indonesia), about 11 years old and "from diverse family backgrounds" from August-December 2022.

Independent Variables: 1) Number of books in the home, 2) number of other learning supports at home, 3) student attitudes toward mathematics, 4) math learning anxiety, & 5) "demographic data" which was identified in the paper tables as "family income"

Dependent Variable: Math achievement (math testing results)



Vientiane
ວຽງຈັນ
Thailand
Bangkok
กรุงเทพมหานคร
Vietnam
Cambodia
Phnom Penh
ភ្នំពេញ
Dalat
Nha Trang
Ho Chi Minh City
Gulf of Thailand
Paracel Islands
South China Sea
Spratly Islands
Palawan
Luzon
Manila
Mindoro
Philippines
Samar Island
Panay
Cebu
Negros
Sulu Sea
Mindanao
Davao
Palau
SABAH
Brunei
Celebes Sea
PERAK
Malaysia
RIAU ISLANDS
SARAWAK
Singapore
RIAU
JAMBI
EAST KALIMANTAN
WEST KALIMANTAN
CENTRAL KALIMANTAN
SOUTH KALIMANTAN
Manado
NORTH MALUKU
NORTH SULAWESI
CENTRAL SULAWESI
WEST SULAWESI
SOUTH SULAWESI
Sorong
WEST PAPUA
BANDA ARC
MALUKU
Indonesia
Makassar
Banda Sea
SOUTH SUMATRA
LAMPUNG
Java Sea
Jakarta
Bandung
Semarang
Surabaya

How variables were measured

- Number of books in the home and learning supports at home: TIMSS 2019 questionnaire (TIMSS is an international math test)
- Student attitudes toward math: Attitudes toward Mathematics Inventory (ATMI) simple version which measures mathematics *motivation*, *self-confidence* in mathematics, perceived *value* of mathematics
- Math learning anxiety: Modified Abbreviated Mathematics Anxiety Scale (mAMAS) questionnaire
- Math achievement: Final Academic Session Examination 2022 (FASE). The mathematics questions in FASE 2022 are obtained through the Instrument Collection and Installation Application (ICIA) system

Overview of Methods

- Researchers used three selection techniques: stepwise, forward addition, and backward elimination to test the contribution of each independent variable.
- Stepwise: test the contribution of each independent variable by sequentially adding variables based on their significance. The variable with the largest contribution is added first, followed by others based on incremental contribution to the model.
- Forward addition: starting with one independent variable and adding others incrementally (similar to stepwise)
- Backward elimination: begins with all independent variables included in the model, and then sequentially removes those that do not contribute significantly
- The Durbin-Watson test was used to check for the presence of autocorrelation (if errors are independent - an assumption of linear regression).
- Researchers tested for multicollinearity problems by checking variance inflation factor (VIF) and tolerance. No multicollinearity problems were found.
- Cook's Distance was used to identify outliers. No outliers were found.

Quick Review: Multiple Regression Assumptions

- homogeneity of variance
- normality of distribution
- linearity

Within the study, variables which did not meet the normality assumption were transformed. Transformed variables (log) were: math learning anxiety & student attitude: value.

Model 1 of 2

- Stepwise, forward, backward selection
- Used all variables without domains (math anxiety: learning & testing; attitudes: motivation, confidence, value)
- May have violated multiple regression assumptions
- Unclear how assumptions may have been violated
- ($p<0.05$) - significant model but unreliable due to bullets above
- Not reporting much on this model because less reliable

Model 2 of 2

- Stepwise, forward, backward selection
- Used variables within domains (Math anxiety: learning & testing; Student attitudes: motivation, confidence, value)
- Focused on fidelity to multiple regression assumptions
- Researchers concluded that the second model was most accurate because it had a lower standard error of estimation, a higher R^2 and a higher adjusted R^2.
- Researchers also noted the importance of fidelity to assumptions
- Model 2 accounted for 27.6% of variance in mathematics achievement. Three variables (out of five variables) were found to have the greatest impact on the dependent variable of mathematics achievement
- The study concluded that three variables (out of five) contribute to student mathematics achievement: number of books in the home (0.315), student confidence (0.224) , and mathematics learning anxiety (-0.256) .
- ($p<0.05$) - significant model

Note on Stepwise Regression (Backward & Forward): Methodological Controversy

- The conceptual goal of these researchers is to explain variance in math achievement.
- The methodological choice of these researchers to use stepwise, backward & forward selection, which is associated with predictive modeling
- Keith, Ch 5, cautions against using stepwise regression as an explanatory method to understand why mathematical achievement varies.
- Shmueli describes explanatory modeling as a way to understand relationships among variables and predictive modeling as a way to forecast unseen observations. Again, within Shmueli's framework, this paper would be seen as explanatory in purpose yet predictive in methodology.
- Mismatch between purpose and methodology.

Reasons for preferring one methodology over another

- For explanatory modeling, it is preferable to use theory-driven models instead of statistical criteria.
- In stepwise selection, variables are added in order of importance, and yet the order is not actually known in explanatory modeling. Keith, on p 106 of Ch 5 in our reading, called this circular reasoning, which is why he does not recommend stepwise selection for explanatory modeling.
- "The reason that stepwise regression does not help in determining the importance of variables is because using R^2 as a measure of the importance of variables is predicated on the assumption that the variables have been entered in the correct order. "

Limitations of the Paper

- Previously discussion of using predictive methodology for an explanatory goal
- Unclear math anxiety labeling. Even when accounting for the differences between models, the distinction between math learning anxiety and math testing anxiety was not consistently drawn. This is a major difference.
- Inconsistent demographic naming within the paper. For example, “demographics” within the paper was labeled “family income” within the tables and described as “diverse family backgrounds” in the abstract.
- Overall, there were several issues with lack of clarity.
- Translation issues? Was the paper originally in English or not?
- It also seems unnecessary to test between the two models. Testing for the necessity of adhering to assumptions seems repetitive, and the lack of domains seems nonsensical. Researchers could have focused on increasing the clarity of Model 2.

Strengths of the Paper

- Addressed a crucial topic (factors affecting student math achievement)
- Provides correlation possibilities between math learning anxiety/student confidence and performance. May result in practical action steps.
- Some demographics provided (number & age of students)
- Emphasis on necessity of following multiple regression assumptions through testing two models, one focusing on fulfilling assumptions and one not
- However, caution should be used when interpreting explanatory claims, since stepwise selection was used - a methodology best suited for prediction

Do these findings about students in Malaysia seem applicable to U.S. students?

- As a refresher, the study concluded that three variables (out of five) contribute to student mathematics achievement: number of books in the home, student confidence, and mathematics learning anxiety. (The non-influential variables were demographics, other learning supports in the home, math testing anxiety, student value of math, and student motivation toward math).
- Student confidence seems repetitive with mathematics learning anxiety (some undetected multicollinearity?).
- I think that the number of books in the home *could* be an influential variable but more likely is indicative of parent achievement & influence; therefore, demographics might be more applicable in the U.S. Again, these demographics seem to have been “family income” but is unclear.
- Math learning anxiety does seem like an influential variable in U.S. students.
- Again, however, I believe math testing anxiety has more influence on math test achievement for U.S. students.
- I am unconvinced of the researchers’ results. There were many instances of having to double-check vocabulary & naming, which added complexity to interpreting the paper in addition to understanding the statistical results and significance. This, in addition to using stepwise, makes me cautious in interpreting their results.

Key Takeaways

- Be consistent when naming variables
- Be clear about sources of information within datasets
- Be clear about the purpose of the model (explanatory or predictive) and use appropriate methods
- Each of the bullets above caused consternation in me when I was interpreting the paper. It was a valuable example of how I can increase accuracy & clarity when utilizing methodologies & reporting results.