Journal of Mental Health & Well-being (JMHWB)

VOLUME 2 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Journal of Mental Health & Well-being (JMHWB)

Volume 2 Issue 1, Year 2026
ISSN: 3142-7510 (Online)
DOI: https://doi.org/10.54536/jmhwb.v2i1.8205
https://journals.e-palli.com/home/index.php/jmhwb

Model Selection Analysis for Predicting Students Mental Health from Academic and Psychological Variables

Ahmed Rageh Ismail[1*]

Article Information

Received: June 20, 2026

Accepted: September 05, 2026

Published: September 23, 2026

Keywords

FWDSelect, Model Selection, Mental Health Index, Multicollinearity, Student Mental Health, Stepwise Regression

ABSTRACT

This paper proposes a complete model selection analysis in order to predict mental health index based on 15 predictor variables taken from student mental health and burnout dataset (N = 2,549). The approach uses the stepwise regression method, which includes forward, backward and bidirectional selection procedures along with the FWDselect method for predictor selection with maximal information content. The results show that the entire 15-predictor model provides perfect goodness of fit ($R^2 = 1.0$), having near-zero residuals (10^{-10}), which means that the mental health index in the given case is entirely defined by the predictor variables as a linear combination of stress level, anxiety score, and depression score. Stepwise regression allows reducing the entire model down to six core predictors (stress level, anxiety score, depression score, sleep hours, dropout risk and study hours per day) without any loss of predictive power. The FWDselect procedure proves stress level as the only necessary predictor, providing nearly 90.7% of variance, while three mental health subscales are selected at the first positions across all model sizes. There is no problem with multicollinearity (all VIF < 10) despite the perfect fit, since the latter is quite high and does not affect the reliability of the model. Thus, one can conclude that the mental health index in the dataset is primarily constructed from psychological distress measures, raising important questions about the construct validity of composite mental health indices and the implications for predictive modelling in mental health research.

INTRODUCTION

Background

The prediction and measurement of psychological well-being in student populations has emerged as a key area of research importance in recent decades (Eisenberg *et al.*, 2007; Auerbach *et al.*, 2016). The increasing trend of psychological distress amongst students at universities around the world has necessitated research into the variables influencing the mental health of individuals, which include academic pressures, financial stress, social support, and personal habits (Hunt & Eisenberg, 2010; Scott *et al.*, 2016). Nonetheless, one of the most basic problems in this field is operationalizing and measuring psychological health outcomes, where researchers tend to use composite measures based on individual indicators.

There exist some issues that are worth considering from the point of view of methodology when it comes to creating indices of mental well-being. If mental well-being is determined using a set of psychological measures such as stress, anxiety, and depression, then the resulting index will have either perfect or close-to-perfect statistical association with the mentioned variables (Streiner *et al.*, 2015). Such an approach may affect the results of any kind of regression analysis in terms of predictive power since one should not include the predictors that mathematically determine the outcome in the model (Babyak, 2004; Harrell, 2015).

Model Selection in Mental Health Research

The step of model selection is crucial to the process of building a predictive model since it consists of selecting a set of predictor variables that are both informative and parsimonious (Burnham & Anderson, 2002). In the area of mental health, the problem of model selection is especially acute because, usually, there are a number of different predictors. The most common approaches in the field of model selection include stepwise regression approaches, namely, forward selection, backward elimination, bidirectional selection and the FWDselect algorithm (Sestelo Perez *et al.*, 2016).

However, it is widely recognized that the stepwise regression procedures have certain weaknesses and are rather controversial approaches to model selection (Harrell, 2015; Steyerberg, 2019). According to Harrell (2015), stepwise regression is not recommended in the practice of statistics because it leads to an unreliable solution, increases Type I error rates and does not solve the problem of multicollinearity between predictors (Derksen & Keselman, 1992; lom & Cassell, 2007). However, if used carefully, the stepwise procedure could be helpful for identifying important variables.

The FWDselect Algorithm

The FWDselect package offers an alternative forward selection approach with various information criteria, providing a more systematic framework for variable selection (Sestelo Perez *et al.*, 2016). This algorithm identifies optimal subsets of predictors of specified sizes using criteria such as the Akaike Information Criterion

[1] Arab Open University, Egypt
[*] Corresponding author's e-mail: a.rageh@rocketmail.com

(AIC) and the corrected Akaike Information Criterion (AICc), which balance model fit against complexity (Burnham & Anderson, 2002). The qselection procedure allows researchers to explore the optimal variable subset for each size, providing a comprehensive picture of the variable selection path (Sestelo Perez *et al.*, 2016).

Study Aims and Significance

This study aims to

- Compare the performance of forward, backward, and bidirectional stepwise regression techniques in selecting predictors for mental health index.
- Apply the FWDselect algorithm to identify optimal variable subsets of varying sizes.
- Assess the impact of multicollinearity on model stability and interpretability.
- Evaluate the implications of using mathematically derived composite outcome variables in predictive modelling.

Understanding the structure of composite mental health indices and the relationships among their component variables is essential for both substantive and methodological reasons. Substantively, this understanding can illuminate which psychological factors are most central to students' mental health experiences. Methodologically, it can guide researchers in selecting appropriate analytical approaches and interpreting results with appropriate caution.

LITERATURE REVIEW

Predictors of Student Mental Health

A number of studies have examined different factors associated with students' mental well-being. Psychological distress indicators including perceived stress, anxiety, and depression were found to be the most significant correlates of overall mental health (Eisenberg *et al.*, 2007). Such concepts are often measured using valid scales like the Perceived Stress Scale (Cohen *et al.*, 1983), the Generalized Anxiety Disorder-7 (Spitzer *et al.*, 2006), and the Patient Health Questionnaire-9 (Kroenke *et al.*, 2001). Academic and lifestyle factors are also associated with mental health in college students. The number of hours devoted to studying, pressure of examinations, and academic success are related to psychological distress, where high requirements for studying may lead to burnout and poor well-being (Schaufeli *et al.*, 2002). Adequate sleep and physical activities are among protective factors, which ensure positive mental health outcomes (Hirshkowitz *et al.*, 2015; Penedo & Dahn, 2005).

There is also a number of social and environmental factors that affect students' mental well-being, including social support, financial stress, and family pressure (Nasr *et al.*, 2024). Effective support from other people allows moderating stress effects, while financial problems and family demands may increase psychological distress. Strong social support networks can buffer the effects of stress, while financial strain and family pressure can exacerbate psychological distress (Cohen & Wills, 1985; Eisenberg *et al.*, 2007).

Multicollinearity and Its Implications

Multicollinearity-the presence of high correlations among predictor variables-represents a significant challenge in regression analysis (Belsley *et al.*, 2005). When predictors are highly correlated, it becomes difficult to estimate the unique contribution of each variable, leading to unstable coefficient estimates, inflated standard errors, and reduced statistical power (Fox, 2015). The Variance Inflation Factor (VIF) is commonly used to assess the severity of multicollinearity, with VIF values exceeding 10 generally considered indicative of problematic multicollinearity (O'Brien, 2007).

In mental health research, multicollinearity is particularly common because psychological constructs such as stress, anxiety, and depression are conceptually related and often highly correlated (Clark *et al.*, 1994). This can complicate model selection, as stepwise procedures may produce different results depending on the order in which variables are entered (Graham, 2003). Understanding the structure of these correlations is essential for both substantive interpretation and methodological decision-making (Steyerberg, 2019).

Model Selection: Methodological Considerations

Model selection in mental health research requires careful consideration of multiple factors (Harrell, 2015). Parsimony-the preference for simpler models with fewer predictors-is a guiding principle, reflecting the belief that explanatory power should not come at the cost of unnecessary complexity (Burnham & Anderson, 2002). Information criteria such as AIC and AICc provide a formal mechanism for balancing fit and complexity, with lower values indicating better models (Akaike, 1974; Sugiura, 1978).

The stability and reliability of selected models is another important consideration (Heinze *et al.*, 2018). Stepwise selection procedures can produce unstable solutions, particularly when data are limited or predictors are highly correlated (Derksen & Keselman, 1992). Cross-validation and bootstrap resampling can help assess the generalisability of selected models (Steyerberg, 2019).

The interpretability of selected models is also crucial (Babyak, 2004). Researchers must consider whether selected variables make theoretical sense and whether the model provides useful insights into the phenomenon of interest. This requires careful attention to the nature and measurement of both outcome and predictor variables (Streiner *et al.*, 2015).

Gaps in the Literature

Despite the extensive literature on student mental health, several gaps remain. First, few studies have systematically examined the structure of composite mental health indices and their relationships with component variables. Second, the implications of using mathematically derived outcome measures in predictive modelling have not been adequately explored. Third, comparative evaluations of different model selection techniques in the context of student mental health data are limited.

This study addresses these gaps by conducting a comprehensive model selection analysis of a mental health index using 15 predictor variables, comparing stepwise methods with the FWDselect algorithm, and examining the implications of multicollinearity and construct measurement for substantive interpretation.

MATERIALS AND METHODS

Data Source and Variables

This study utilized the "student_mental_health_burnout.csv" dataset, accessed from Kaggle (Sharmaji, 2024), comprising 2,549 observations of university students. The dataset includes measures of mental health and burnout.

Dependent Variable

- mental_health_index — a composite continuous score ranging from 2.027 to 10.000, with higher values indicating better mental health.

Independent Variables (15 predictors):

- study_hours_per_day — Daily study hours
- exam_pressure — Perceived exam pressure
- academic_performance — Academic performance score
- stress_level — Self-reported stress level
- anxiety_score — Anxiety assessment score
- depression_score — Depression assessment score
- sleep_hours — Average daily sleep
- physical_activity — Physical activity level
- social_support — Social support network score
- screen_time — Daily screen time hours
- internet_usage — Internet usage intensity
- financial_stress — Financial stress level
- family_expectation — Family pressure
- burnout_score — Burnout assessment
- dropout_risk — Predicted dropout risk

Analytical Approach

The analysis proceeded through the following stages:

Stage 1: Correlation Matrix Analysis. Pearson correlation coefficients were computed among all variables to identify relationships with mental health index and detect multicollinearity among predictors.

Stage 2: Full Model Regression. An Ordinary Least Squares (OLS) regression was estimated with all 15 predictors to establish the baseline full model and assess multicollinearity via Variance Inflation Factors (VIF) (Fox, 2015; O'Brien, 2007).

Stage 3: Stepwise Regression. Three stepwise selection procedures were applied (Heinze *et al.*, 2018):

- **Forward Selection:** Starting from a null model, predictors were sequentially added based on AIC.
- **Backward Elimination:** Starting from the full model, predictors were sequentially removed based on AIC.
- **Bidirectional (Stepwise) Selection:** Combining forward and backward steps, allowing variables to enter and exit.

All stepwise procedures used AIC as the selection criterion (Akaike, 1974).

Stage 4: FWDselect Algorithm. The FWDselect package was employed to identify optimal variable subsets of specified sizes using:

- **Best Single Predictor (q = 1):** Identified using R^2 criterion.
- **Best 8-Variable Model (q = 8):** Identified using AICc criterion (Sugiura, 1978).
- **qselection Procedure:** Optimal subsets identified for each size (q = 1 to 9) using deviance and AICc criteria (Sestelo Perez *et al.*, 2016).

RESULTS DISCUSSION

Correlation Matrix Analysis

This shows that there is high correlation between stress, anxiety, depression and burnout and mental well-being index because these variables seem to form mathematical formula for the construction of mental well-being index.

Table 1: Correlations with Mental Health Index

Variable	Correlation
stress_level	-0.953
anxiety_score	-0.873
depression_score	-0.753
burnout_score	-0.804
dropout_risk	-0.648
financial_stress	-0.457

High correlations among distress measures suggest potential multicollinearity issues, while the near-perfect

Table 2: Predictor Multicollinearity

Variable Pair	Correlation
stress_level ↔ anxiety_score	0.776
stress_level ↔ burnout_score	0.760
anxiety_score ↔ burnout_score	0.678
screen_time ↔ internet_usage	0.887

correlation between screen_time and internet_usage may warrant consideration.

Full Model Regression
Model Summary:
Call: lm(formula = mental_health_index ~ ., data = MyData[,])
Residuals:
Min 1Q Median 3Q Max
-8.73e-10 -2.65e-10 4.20e-12 2.54e-10 8.98e-10
Multiple R-squared: 1, Adjusted R-squared: 1
F-statistic: 2.689e+21 on 15 and 2533 DF, p-value: < 2.2e-16

All VIF values are below 10 (O'Brien, 2007), indicating that multicollinearity is not severe enough to invalidate the model despite the perfect fit.

Table 3: Coefficient Estimates

Predictor	Estimate	Std. Error	t value	Significance
(Intercept)	10.000	1.07e-10	9.35e+10	***
stress_level	-0.400	9.27e-12	-4.32e+10	***
anxiety_score	-0.300	7.04e-12	-4.26e+10	***
depression_score	-0.300	8.30e-12	-3.62e+10	***
study_hours_per_day	-9.78e-12	5.75e-12	-1.701	.
dropout_risk	1.67e-11	7.93e-12	2.099	*
All other predictors	≈ 0	—	≈ 0	Not significant

Table 4: Variance Inflation Factors (VIF)

Variable	VIF
stress_level	5.596
internet_usage	4.728
screen_time	4.725
burnout_score	4.240
exam_pressure	3.481
study_hours_per_day	3.181
anxiety_score	2.727
depression_score	2.390

Stepwise Regression Results

Forward Selection: No variables were added because the full model already had the minimum AIC. Start AIC = -111,261.6. The full model retained all 15 predictors.

Final Backward Model (6 predictors)

mental_health_index ~ study_hours_per_day + stress_level + anxiety_score + depression_score + sleep_hours + dropout_risk

Table 5: Backward Elimination

Step	Variable Removed	ΔAIC	Variables Remaining
1	screen_time	-2.5	14
2	internet_usage	-2.0	13
3	physical_activity	-1.8	12
4	exam_pressure	-1.7	11
5	academic_performance	-1.5	10
6	burnout_score	-1.5	9
7	family_expectation	-1.5	8
8	financial_stress	-1.7	7
9	social_support	-1.3	6

Table 6: Final Model Coefficients

Predictor	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.000	4.07e-11	2.46e+11	<2e-16 ***
stress_level	-0.400	7.34e-12	-5.45e+10	<2e-16 ***
anxiety_score	-0.300	6.80e-12	-4.41e+10	<2e-16 ***

depression_score	-0.300	7.76e-12	-3.87e+10	<2e-16 ***
study_hours_per_day	-5.59e-12	3.44e-12	-1.623	0.1047
sleep_hours	7.02e-12	4.68e-12	1.500	0.1336
dropout_risk	1.66e-11	6.68e-12	2.484	0.0131 *

Model Statistics

- R^2 = 1.0 (perfect fit maintained)
- **Residual standard error:** 3.31e-10
- **F-statistic:** 6.735e+21 on 6 and 2542 DF
- **AIC:** -111,276

Bidirectional (Stepwise) Selection

Produced the identical final model as backward elimination, confirming robustness.

AIC improved (became more negative) with each removal, indicating that simpler models are preferred when all models fit perfectly.

Table 7: Stepwise Selection Path

Step	Action	Variables Removed	AIC
1	Initial	—	-111,262
2	Remove	screen_time	-111,264
3	Remove	internet_usage	-111,266
4	Remove	physical_activity	-111,267
5	Remove	exam_pressure	-111,269
6	Remove	academic_performance	-111,271
7	Remove	burnout_score	-111,272
8	Remove	family_expectation	-111,274
9	Remove	financial_stress	-111,275
10	Remove	social_support	-111,276

FWDselect Algorithm Results

Best Single Predictor (q = 1, criterion = R^2)

- **Best variable:** stress_level
- **Coefficients**: (Intercept) = 10.221, stress_level = -0.750
- **Null deviance:** 4429
- **Residual deviance:** 409.7
- **AIC:** 2580

When forced to choose one predictor for mental health index, stress_level alone explains approximately 90.7% of the variance (residual deviance reduced from 4429 to 409.7).

Best 8-Variable Model (q = 8, criterion = AICc)

Selected variables

- x[4] = stress_level
- x[5] = anxiety_score
- x[6] = depression_score
- x[15] = dropout_risk
- x[1] = study_hours_per_day
- x[7] = sleep_hours
- x[9] = social_support
- x[3] = academic_performance

Model Fit

- **Residual deviance:** 2.78e-16 (essentially zero)
- **AIC:** -104,038 (extremely low, indicating perfect fit)
- **Dispersion parameter:** 1.10e-19

Table 8: Variable Selection Path (qselection with deviance criterion)

Subset Size	Variables Selected (in order of addition)
1	stress_level
2	stress_level, anxiety_score
3	stress_level, anxiety_score, depression_score
4	+ dropout_risk
5	+ study_hours_per_day
6	+ sleep_hours
7	+ social_support
8	+ academic_performance
9	+ exam_pressure

qselection with AICc Criterion: Produced an identical selection path, confirming the robustness of the variable ordering.

The three mental health subscales (stress_level, anxiety_score, depression_score) are consistently selected first, confirming they are the primary determinants of mental health index. This matches the perfect fit pattern observed in the regression models.

Discussion

Interpretation of Findings

Perfect Fit and Its Implications

The most striking finding of this analysis is the perfect fit achieved by both the full model ($R^2 = 1.0$) and the reduced 6-predictor model ($R^2 = 1.0$), with residuals approaching zero (10^{-10} magnitude). This level of fit is exceptionally rare in real-world data and indicates that mental_health_index is not merely associated with the predictors but is mathematically derived from them—specifically, as a linear combination of stress_level, anxiety_score, and depression_score (Streiner *et al.*, 2015).

The coefficient estimates provide strong evidence for this interpretation. The three psychological distress measures—stress_level (-0.400), anxiety_score (-0.300), and depression_score (-0.300)—account for the entire variation in mental health index, with all other predictors having coefficients effectively equal to zero. This suggests that mental_health_index is constructed as a weighted sum of these three components, rather than being independently measured (Clark *et al.*, 1994).

This finding has significant implications for both substantive interpretation and methodological practice. Substantively, it suggests that the composite mental health index does not capture aspects of mental health beyond stress, anxiety, and depression. While these are important dimensions of psychological wellbeing, a composite index that is mathematically determined by its subscales cannot provide additional information beyond those component measures (Streiner *et al.*, 2015). Researchers using such indices should be aware that they are essentially reusing the same information, potentially inflating apparent relationships with other variables.

Reduced Model and Variable Selection

The backward elimination procedure reduced the full 15-predictor model to six core variables (stress_level, anxiety_score, depression_score, sleep_hours, dropout_risk, and study_hours_per_day) without any loss of explanatory power. The retention of mental health index is primarily determined by psychological distress, with the core three variables (stress, anxiety, depression) accounting for the majority of the variance. The inclusion of sleep_hours, dropout_risk, and study_hours_per_day appears to be due to their correlations with the primary distress measures rather than independent contributions.

The bidirectional selection procedure produced an identical final model, providing some reassurance about the stability of the selection process (Heinze *et al.*, 2018). However, the perfect fit and near-zero residual standard errors render traditional model selection criteria such as AIC less informative, as all models fit perfectly. In such cases, AIC tends to favour simpler models, as observed with the improvement in AIC with each variable removal.

FWDselect Confirmation

The FWDselect algorithm confirmed and extended the stepwise findings. The best single predictor was stress_level, explaining approximately 90.7% of the variance in mental health index—a remarkably high figure that underscores the centrality of stress in the composite measure (Cohen *et al.*, 1983). The variable selection path consistently selected stress_level first, followed by anxiety_score and depression_score, with other variables added only after these core predictors were included.

The consistency between stepwise regression and FWDselect results provides strong evidence that the core predictors of mental health index are the psychological distress measures (Sestelo Perez *et al.*, 2016). This consistency also demonstrates the robustness of the variable ordering across different selection approaches and criteria.

Multicollinearity Assessment

The VIF values obtained in this analysis are somewhat unusual given the perfect fit. While all VIF values are below the conventional threshold of 10 (O'Brien, 2007), indicating that multicollinearity is not severe enough to invalidate the model, the perfect fit and near-zero coefficient standard errors complicate the interpretation. The VIF values for stress_level (5.60), internet_usage (4.73), screen_time (4.72), and burnout_score (4.24) indicate moderate correlations with other predictors. Given the perfect fit, these VIF values likely reflect the deterministic relationships between the outcome and its component variables rather than typical multicollinearity (Belsley *et al.*, 2005).

Contributions to the Literature

This study makes several contributions to the literature on student mental health and predictive modelling. First, it provides a systematic demonstration of how composite outcome variables can influence model selection and interpretation. The finding that mental_health_index is mathematically derived from psychological distress measures highlights the importance of understanding the construction of outcome variables (Streiner *et al.*, 2015). Second, the study provides a comparative evaluation of stepwise regression and FWDselect algorithms in a mental health context, offering guidance for researchers facing similar analytical challenges (Sestelo Perez *et al.*, 2016; Heinze *et al.*, 2018). Third, the findings illuminate the central role of stress, anxiety, and depression in shaping overall mental health, consistent with theoretical frameworks emphasising the interconnections among these constructs (Clark *et al.*, 1994).

Limitations and Future Directions

Several limitations should be acknowledged. First, the perfect fit and deterministic relationships in this dataset indicate that mental_health_index is a mathematically derived composite, limiting the generalisability of the findings to contexts where mental health is measured independently (Streiner *et al.*, 2015). Second, the analysis is based on a single dataset, and the findings should be replicated in other student mental health datasets. Third, the cross-sectional nature of the data precludes causal inference; longitudinal designs are needed to establish temporal relationships.

Future research should explore the application of model selection techniques to independently measured mental health outcomes, where moderate rather than perfect fit is expected. Additionally, researchers should explicitly report the construction of composite variables and consider the implications of using mathematically derived outcomes in predictive models (Babyak, 2004). The development and validation of mental health measures that capture a broader range of wellbeing dimensions, beyond stress, anxiety, and depression, would also be valuable.

CONCLUSION

This study demonstrates that the mental health index in the student dataset is primarily determined by psychological distress measures, specifically stress_level, anxiety_score, and depression_score. The perfect fit achieved by regression models ($R^2 = 1.0$) indicates that mental_health_index is mathematically derived from these component variables, with all other predictors having negligible independent contributions. Stepwise regression reduces the full 15-predictor model to six core variables without loss of explanatory power, while the FWDselect algorithm confirms stress_level as the single best predictor.

The practical implications of these findings are significant. Researchers using composite mental health indices should be aware that such measures may not capture aspects of mental health beyond their component subscales (Streiner *et al.*, 2015). If the index is a linear combination of stress, anxiety, and depression, it cannot provide additional information beyond what those separate measures already provide. This has implications for study design, variable selection, and the interpretation of regression coefficients.

Methodologically, this study underscores the importance of understanding the construction of outcome variables before conducting model selection (Harrell, 2015). When outcomes are mathematically derived from predictors, traditional model selection procedures may produce perfect fit and misleading conclusions about variable importance. Researchers should routinely examine the relationships among outcome and predictor variables and consider the implications of deterministic relationships for statistical inference (Babyak, 2004).

In conclusion, the mental health index in this dataset is effectively a weighted sum of stress, anxiety, and depression, offering little beyond these constituent measures. By understanding the mathematical structure of composite indices, researchers can make more informed decisions about variable selection, model interpretation, and the substantive conclusions drawn from their analyses.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., Hwang, I., Kessler, R. C., Liu, H., Mortier, P., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., Benjet, C., Caldas-de-Almeida, J. M., Demyttenaere, K., ... Bruffaerts, R. (2016). Mental disorders among college students in the World Health Organization World Mental Health Surveys. *Psychological Medicine, 46*(14), 2955–2970.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Biopsychosocial Science and Medicine, 66*(3), 411-421.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*(1), 103–116.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98*(2), 310–357.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology, 45*(2), 265–282.

Eisenberg, D., Gollust, S. E., Golberstein, E., & Hefner, J. L. (2007). Prevalence and correlates of depression, anxiety, and suicidality among university students. *American Journal of Orthopsychiatry, 77*(4), 534–542.

Flom, P. L., & Cassell, D. L. (2007). *Stopping stepwise: Why stepwise selection is bad and what you should use instead.* Proceedings of the SAS Global Forum 2007 Conference. https://www.lexjansen.com/nesug/nesug07/cc/CC09.pdf

Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd ed.). SAGE Publications.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology, 84*(11), 2809–2815.

Harrell, F. E. (2015). *Regression modeling strategies: With* *applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer.

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection—A review and recommendations for the practicing statistician. *Biometrical Journal, 60*(3), 431–449.

Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M. V., Ware, J. C., & Adams Hillard, P. J. (2015). National Sleep Foundation's sleep time duration recommendations: Methodology and results summary. *Sleep Health, 1(*1), 40–43.

Hunt, J., & Eisenberg, D. (2010). Mental health problems and help-seeking behavior among college students. *Journal of Adolescent Health, 46*(1), 3–10.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Nasr, R., Abdel Rahman, A., Haddad, C., Nasr, N., Karam, J., Hayek, J., Ismael, I., Swaidan, E., Salameh, P., & Alami, N. (2024). The impact of financial stress on student wellbeing in Lebanese higher education. *BMC Public Health, 24*(1), Article 1234.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690.

Penedo, F. J., & Dahn, J. R. (2005). Exercise and well-being: A review of mental and physical health benefits associated with physical activity. *Current Opinion in Psychiatry, 18*(2), 189–193.

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Sestelo Perez, M., Martinez Villanueva, N., Meira Machado, L., & Roca Pardiñas, J. (2016). *FWDselect: An R package for variable selection in regression models*. R Journal.

SSchaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*(1), 71–92.

Scott, K. M., Lim, C., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., Caldas-de-Almeida, J. M., ... & Kessler, R. C. (2016). Association of mental disorders with subsequent chronic physical conditions: world mental health surveys from 17 countries. *JAMA psychiatry, 73*(2), 150-158.

Sharmaji. (2024). *Student mental health and burnout* [Data set]. Kaggle. https://www.kaggle.com/datasets/sharmajicoder/student-mental-health-and-burnout/data

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097.

Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2nd ed.). Springer.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods, 7*(1), 13–26.