



Applying Principal Component Regression to Factors Associated with College Students' Frequency of Cooking Their Own Breakfast

Lale Midad Haqiqia¹, Bunga Shafa Kireina², Aurellia Putri Magani Trihapsari³, Firstia Nabila Ariyani⁴

¹Data Science Program Study of Universitas Pembangunan Nasional "Veteran" Jawa Timur, ²Data Science Program Study of Universitas Pembangunan Nasional "Veteran" Jawa Timur, ³Data Science Program Study of Universitas Pembangunan Nasional "Veteran" Jawa Timur, ⁴Robotics and Artificial Intelligence Program Study of Universitas Airlangga

¹25083010072@student.upnjatim.ac.id, ²25083010071@student.upnjatim.ac.id, ³25083010064@student.upnjatim.ac.id,

⁴firstia.nabila.ariyani-2025@ftmm.unair.ac.id

ABSTRACT

Cooking breakfast independently is associated with healthier eating habits, but its frequency among college. Several interrelated factors influence students. This study aimed to model the weekly frequency of cooking breakfast independently (Y) using principal component regression (PCR) with four predictors: confidence in cooking ability (X1), frequency of not cooking because food ingredients had run out (X2), frequency of cooking because food outside the home was more expensive (X3), and frequency of preparing breakfast to take as a packed meal (X4). The four predictors were selected from 15 candidate questions based on conceptual relevance and examined using Spearman's correlation across 12 numerical indicators; X1 and X3 were significantly correlated with Y, while X2 and X4 were retained despite not being significant. The data came from 150 college students who completed an online form, with no missing values. The analysis included descriptive statistics, Pearson's correlation, standardisation, principal component analysis (PCA), and PCR, with the number of components selected through 5-fold cross-validation based on the minimum RMSE. Two principal components explained 61.01% of the variance in the predictors and produced a cross-validation RMSE of 1.5882. The model was jointly significant ($F = 18.65$; $p = 6.03 \times 10^{-8}$), with an R^2 of 0.2024, an adjusted R^2 of 0.1916, and a cross-validated R^2 of 0.1556. On the original scale, X1, X3, and X4 were positively associated with Y, whereas X2 was negatively associated with Y. The model's explanatory power was low to moderate; therefore, interpret the results as associations rather than causal relationships.

Keywords: principal component regression; principal component analysis; breakfast; college students; cross-validation.

I. INTRODUCTION

Many college students skip breakfast as they begin managing their own schedules and food intake. A study of medical students in Karachi reported that 18.6% of respondents skipped breakfast because they lacked time. This factor was most frequently associated with this behaviour (56.3%) [1]. Among students in Ghana's Western Region, only 35.7% ate breakfast regularly; insufficient funds, limited time or class schedules, and the cost of food on campus were reported as barriers [2]. These findings indicate that students' breakfast habits are not solely a matter of willingness but are also influenced by time and financial constraints.

Cooking for oneself is often viewed as one way to maintain diet quality. Among 1,108 first-year students at eight universities, cooking more frequently, having better cooking skills, and planning meals were associated with higher fruit and vegetable intake and a lower body mass index [3]. In Indonesia, a study of 110 students at Diponegoro University showed that how students obtained their food was associated with their eating habits [4]. Therefore, the frequency with which students prepare their own breakfasts merits investigation as a behaviour related to students' dietary patterns.

Cooking frequency is associated with several factors. Food-skills courses that include practical kitchen sessions have been linked to increased cooking confidence and cooking frequency among students [5], while a lack of knowledge about preparing food independently and limited kitchen space and equipment have been reported as barriers [6]. In addition to ability, behavioural questionnaires can measure the availability of food ingredients at home, the cost of eating out, and the habit of preparing food in advance, such as making breakfast to take along [2], [3]. However, questionnaires generally

* Corresponding author.

E-mail address: 25083010072@student.upnjatim.ac.id
<http://doi.org/10.33005/jasid.v2i2.88>

include many questions, not all of which relate to the behaviour being studied. Therefore, predictors need to be selected rationally before modelling; in this study, there were 15 candidate predictor questions.

Because questionnaire score predictors may be correlated with one another, multiple linear regression can potentially produce unstable coefficients. Principal component analysis (PCA) transforms correlated variables into uncorrelated components ordered by decreasing variance [7], and principal component regression. The principal component regression (PCR) approach regresses the response variable on a number of these components, with the number of components typically selected through cross-validation [8]. PCR has been applied to health data in Indonesia to address multicollinearity [9]. Therefore, this study aims to apply PCR to model the frequency with which college students cook their own breakfast using four predictors selected from 15 candidate questions, determine the number of components through 5-fold cross-validation, evaluate the model's performance, and interpret the coefficients on the original variable scale.

II. METHOD

A. Data and Respondents

The data were sourced from the `data_mentah_eda_final.xlsx` file, which contains responses from 150 student respondents. They completed an online form about breakfast-cooking habits. The questionnaire consisted of 16 behavioural questions: one response question and 15 candidate predictor questions. A total of 131 respondents (87.3%) were from UPN "Veteran" East Java. While 19 respondents were from other higher-education institutions, the respondents were distributed across eight faculties. When asked who usually cooks breakfast at home, 87 respondents said their parents, 48 said they cook themselves, 7 said a domestic helper, and 8 reported that they do not cook. The main reasons for cooking breakfast at home were that it is more economical (91 respondents), suits their preferences (33), and is more hygienic (26).

B. Research Variables and Predictor Selection

The response variable (Y) is the number of days per week respondents cook their own breakfast (values range from 1 to 7). Of the 15 candidate questions, three were nominal variables (who usually cooks breakfast, the breakfast-time range, and the main reason for cooking), so they could not be used directly in PCA, which operates on numerical variables. The remaining 12 numerical questions are hereafter referred to as indicators I1–I12 and were all treated as numerical data.

We examined the relationship between each indicator and Y using Spearman's rank correlation because the response scales were discrete [10], at a significance level of 0.05. We categorised the strength of the relationship by the absolute value of the coefficient: less than 0.20, very weak; 0.20 to less than 0.40, weak; 0.40 to less than 0.60, moderate; 0.60 to less than 0.80, strong; and 0.80 or higher, very strong. These categories vary by field [10]. Four predictors were selected based on their conceptual relevance to the research objective: confidence in cooking ability (X1 = I5, ranging from 1 to 5), frequency of not cooking because food ingredients have run out (X2 = I6, ranging from 0 to 7), frequency of cooking because eating out is more expensive (X3 = I9, ranging from 1 to 7), and frequency of preparing breakfast as a packed meal (X4 = I11, ranging from 1 to 7). We used the Spearman correlation results as supporting evidence rather than the sole basis for selection. We excluded the other eight indicators because they measured the cooking process or outcomes (I1, I2, I3, and I12), overlapped with cost or ingredient availability already represented by I3, X2, and X3 (I7, I8, and I10), or were unrelated to Y (I4).

C. Research Process

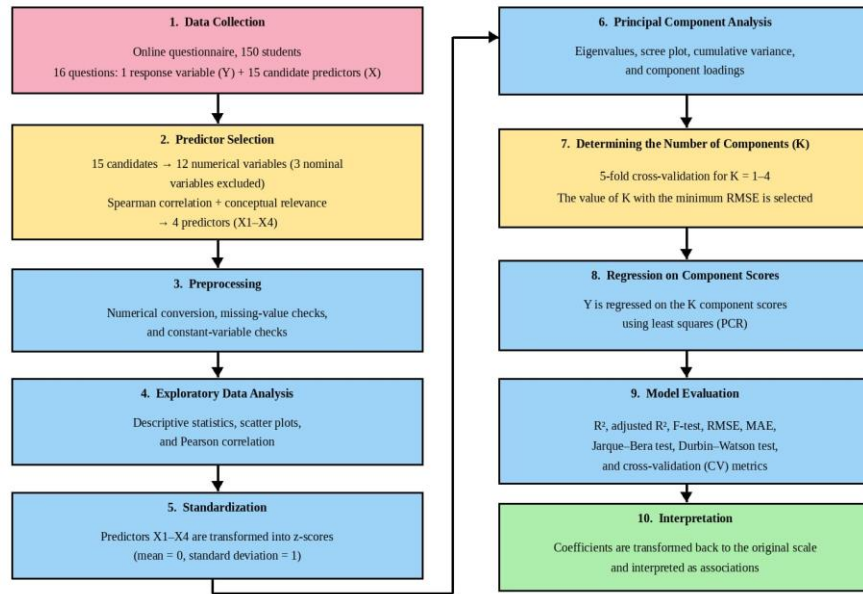


Figure 1. Flowchart of the principal component regression (PCR) study.

Figure 1 shows the research workflow, which consists of ten stages. Questionnaire data from 150 college students were collected first, after which 15 candidate questions were screened down to 12 numerical indicators and ultimately four predictors (X1–X4) through Spearman correlation analysis and conceptual considerations. After data preprocessing and exploration, we standardised the four predictors and used PCA to obtain eigenvalues, cumulative variance, and component loadings. We determined the number of components, K , through 5-fold cross-validation by selecting the smallest RMSE; we then regressed Y on the K component scores to construct a PCR model and evaluated it using goodness-of-fit measures, a simultaneous test, and residual diagnostics. The final stage transformed the component coefficients back to the original variable scale so they could be interpreted as associations.

D. Data Preprocessing and Exploration

The column values were converted to numeric format and checked for completeness; no missing values were found in X1–X4 or Y , so all 150 rows were used. Two rows contained the same combination of values; they were not removed because identical responses can occur naturally on a discrete scale. None of the predictors was constant. The exploration included descriptive statistics (including skewness and kurtosis), scatterplots of X against Y , and Pearson correlations among the variables.

E. Principal Component Analysis

The predictors were first standardised according to Equation (1) so that each variable had a mean of zero and a standard deviation of one. PCA was then applied to the standardised predictors, and the number of components was determined using three criteria: eigenvalues greater than one (Kaiser’s criterion), a minimum cumulative variance of 80%, and the scree plot [7]. Component loadings were used to interpret the variables that formed each component.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

F. Principal Component Regression and Validation

The PCR model in Equation (2) regresses Y on K principal component scores using the least-squares method. We compare $K = 1$ to 4 components through 5-fold cross-validation (shuffling with random state 42) using a standardisation–PCA–linear regression pipeline, with RMSE, MAE, and R^2 as evaluation metrics; we select the K with the smallest cross-validation RMSE. We estimate the final model using statsmodels and evaluate it using R^2 , adjusted R^2 , RMSE, MAE, and the F-test; we check residual assumptions using the Jarque–Bera test and the Durbin–Watson statistic. The component coefficients are transformed to the scale of the original variables using Equations (3) and (4), where a_{ij} is the loading of variable j on component k . All calculations are performed in Python (pandas, SciPy, scikit-learn, and statsmodels).

$$Y_i = \gamma_0 + \gamma_1 PC_{i1} + \dots + \gamma_K PC_{iK} + \epsilon_i \quad (2)$$

$$\beta_j = (1 / s_j) \sum_{k=1}^K a_{jk} \gamma_k \quad (3)$$

$$\beta_0 = \gamma_0 - \sum_j \beta_j \bar{x}_j \quad (4)$$

III. RESULTS AND DISCUSSION

A. Predictor Selection Using Spearman Correlation

Table I shows that, of the 12 numerical indicators, only three were significantly correlated with Y: I9 (cooking because eating out is more expensive; $\rho = 0.3987$; $p < 0.001$), I5 (confidence in cooking; $\rho = 0.2370$; $p = 0.003$), and I12 (number of menu items; $\rho = 0.2051$; $p = 0.012$). All three correlations were positive and weak, whereas the other nine indicators showed very weak and nonsignificant correlations. Among the four selected predictors, X1 and X3 were significant, while X2 ($\rho = -0.1292$; $p = 0.115$) and X4 ($\rho = 0.1354$; $p = 0.099$) were not significant but were retained because they represented the aspects of ingredient availability and packed-meal habits that were the focus of the study. I12 was not used because it described the outcome of cooking activity rather than its determinant, meaning that the direction of its relationship could be reversed. Morning time constraints (I4) did not correlate with Y ($\rho = -0.0480$; $p = 0.560$), even though the literature reports them as a barrier to eating breakfast [1], [2].

TABLE I. SPEARMAN CORRELATIONS BETWEEN 12 NUMERICAL INDICATORS AND Y (N = 150)

Code	Indicator	ρ	p	Strength	Predictor
I1	Time when breakfast preparation begins	-0,0905	0,271	Very Weak	–
I2	Minutes spent cooking breakfast	0,1050	0,201	Very Weak	–
I3	Cost per breakfast preparation	-0,0678	0,410	Very Weak	–
I4	Impact of limited time in the morning	-0,0480	0,560	Very Weak	–
I5	Confidence in cooking ability	0,2370	0,003	Weak	X1
I6	Not cooking because ingredients ran out (times/week)	-0,1292	0,115	Very Weak	X2
I7	Weekly grocery cost	0,0723	0,379	Very Weak	–
I8	Number of days breakfast ingredients last	0,1480	0,071	Very Weak	–
I9	Cooking because eating out is more expensive (times/week)	0,3987	< 0,001	Weak	X3
I10	Amount that can be saved	0,0649	0,430	Very Weak	–
I11	Preparing breakfast to take along (times/week)	0,1354	0,099	Very Weak	X4
I12	Number of different menu items per week	0,2051	0,012	Weak	–

B. Descriptive Statistics

Table II shows that the average frequency of cooking one's own breakfast (Y) was 3.14 times per week, with a median of 3, and only 21.3% of respondents (32 out of 150) cooked for themselves five or more times. The distribution of Y was positively skewed (0.6345), indicating that it was concentrated toward lower frequencies. Confidence in cooking (X1) averaged 3.65 on a 1–5 scale and was negatively skewed toward higher values (–0.4128), while the average frequency of not cooking because ingredients had run out (X2) was 2.37 times per week. Figure 2 presents the distributions of X1–X4 and Y as box plots, with a median of 4 for X1 and medians of 2, 4, 3, and 3.

TABLE II. DESCRIPTIVE STATISTICS OF THE RESEARCH VARIABLES (N = 150)

Variable	Description	Min	Max	Average	Standard Deviation	Skewness	Kurtosis
X1	Confidence in cooking ability (scale of 1–5)	1	5	3,6467	1,0500	–0,4128	–0,1969
X2	Doesn't cook because ingredients have run out (times/week)	0	7	2,3667	1,6726	0,5627	–0,1702
X3	Cooks because eating out is more expensive (times/week)	1	7	3,8533	1,7432	0,2436	–0,7332
X4	Prepares breakfast to take along (times/week)	1	7	2,8667	1,6165	0,3448	–1,1491
Y	Cooks breakfast themselves (times/week)	1	7	3,1400	1,7341	0,6345	–0,3401

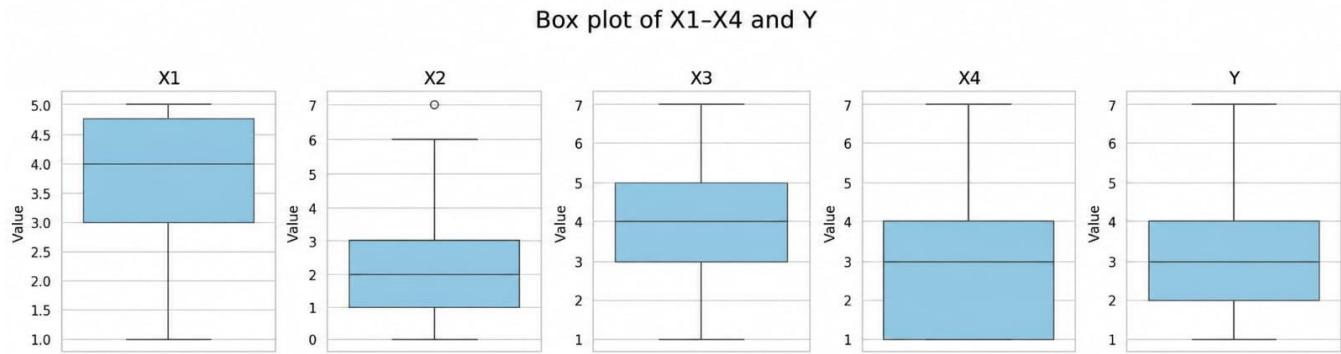


Figure 2. Box plots of X1–X4 and Y.

Figure 3 shows upward trend lines for X1, X3, and X4 and a downward trend line for X2 relative to Y, with points widely scattered around the lines, indicating that no single predictor predominantly explains Y. The points in the figure have been given a small amount of random jitter solely to display overlapping points in the discrete data.

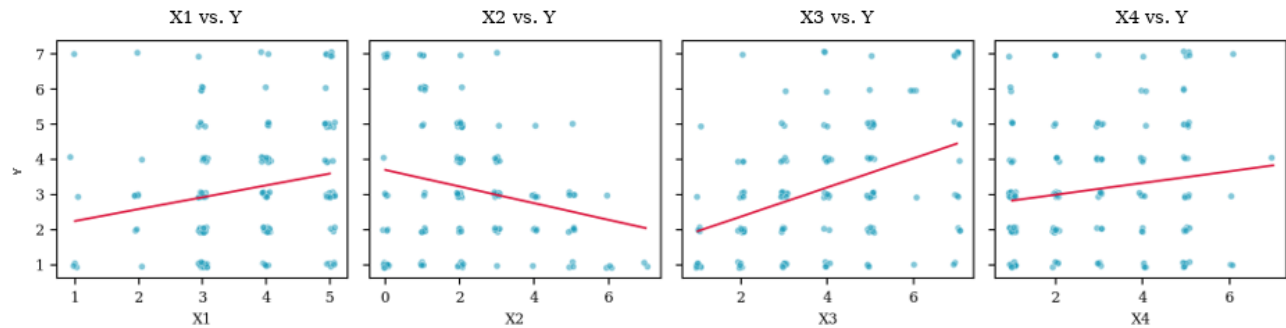


Figure 3. Scatterplot of X1–X4 versus Y with a linear trendline.

C. Correlation Among Variables

Table III shows that Y has the strongest correlation with X3 ($r = 0.4176$), followed by a negative correlation with X2 (-0.2284), X1 (0.2043), and X4 (0.1551). The correlations among the predictors are relatively low; the largest correlation is between X3 and X4 (0.2741), followed by X2 and X3 (-0.2116). These findings indicate that the linear relationships among the predictors are relatively weak. Therefore, this study discusses PCR as a dimensionality-reduction approach, with the number of components selected based on cross-validation.

TABLE III. PEARSON CORRELATIONS

	X1	X2	X3	X4	Y
X1	1,0000	−0,0977	0,0558	−0,0161	0,2043
X2	−0,0977	1,0000	−0,2116	−0,0488	−0,2284
X3	0,0558	−0,2116	1,0000	0,2741	0,4176
X4	−0,0161	−0,0488	0,2741	1,0000	0,1551
Y	0,2043	−0,2284	0,4176	0,1551	1,0000

D. Principal Component Analysis

Table IV shows eigenvalues of 1.3951, 1.0615, 0.8898, and 0.6804. PC1 explains 34.65% of the variance and PC2 explains 26.36%, bringing the cumulative variance explained by the first two components to 61.01%, while three components account for 83.10%. The Kaiser criterion suggests retaining two components because only two eigenvalues exceed one, whereas the 80% threshold is reached only with three components, as shown in the scree plot and variance plot in Figure 4.

TABLE IV. EIGENVALUES AND EXPLAINED VARIANCE

Component	Eigenvalue	Variance (%)	Cumulative (%)
PC1	1,3951	34,65	34,65
PC2	1,0615	26,36	61,01
PC3	0,8898	22,10	83,10
PC4	0,6804	16,90	100,00

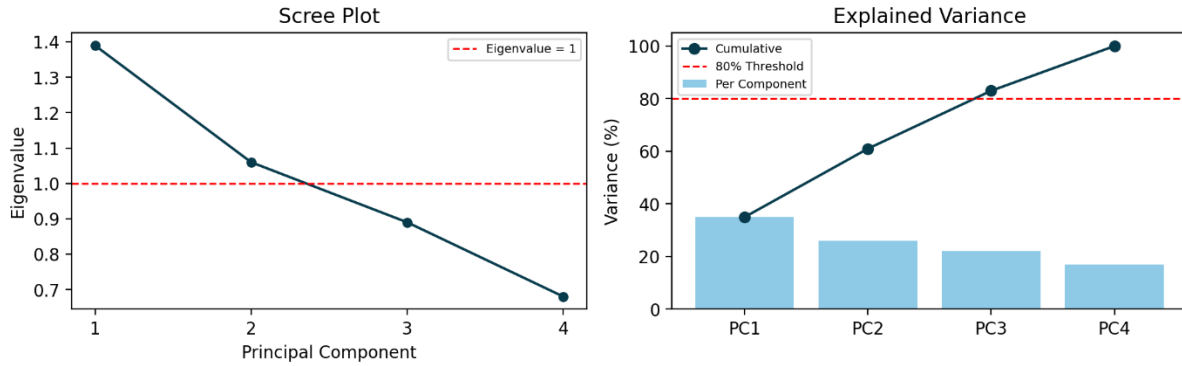


Figure 4. Scree plot (left) and variance explained by each component, along with the cumulative variance (right).

According to Table V, PC1 is primarily formed by X3 (0.6692) and X4 (0.5283), with X2 loading in the opposite direction (-0.4838). It can therefore be interpreted as a dimension of motivation to cook driven by price and supplies, rather than ingredient constraints. PC2 is dominated by X1 (0.7537), with X4 (-0.4965) and X2 (-0.4113) loading in the opposite direction, and can therefore be interpreted primarily as a dimension of cooking confidence.

TABLE V. PRINCIPAL COMPONENT LOADINGS

Variable	PC1	PC2	PC3	PC4
X1	0,1973	0,7537	0,6256	$-0,0391$
X2	$-0,4838$	$-0,4113$	0,6719	0,3812
X3	0,6692	$-0,1277$	$-0,0115$	0,7319
X4	0,5283	$-0,4965$	0,3962	$-0,5634$

E. Selecting the Number of Principal Components

The 5-fold cross-validation in Table VI shows that two components achieved the smallest RMSE (1.5882), with an MAE of 1.2682 and an R^2 of 0.1556. The RMSE difference between one and three components was very small (less than 0.007), so the choice among the three was not practically significant. Four components, equivalent to multiple linear regression using all predictors, produced a higher RMSE (1.6280) and a lower cross-validation R^2 (0.1127). Both the Kaiser criterion and cross-validation indicated two components; therefore, the final model used $K = 2$ (Figure 5).

TABLE VI. 5-FOLD CROSS-VALIDATION PERFORMANCE BY NUMBER OF COMPONENTS

Number of PC	RMSE CV	MAE CV	R^2 CV	Cumulative Variance (%)
1	1,5931	1,2777	0,1503	34,65
2	1,5882	1,2682	0,1556	61,01
3	1,5947	1,2712	0,1487	83,10
4	1,6280	1,2875	0,1127	100,00

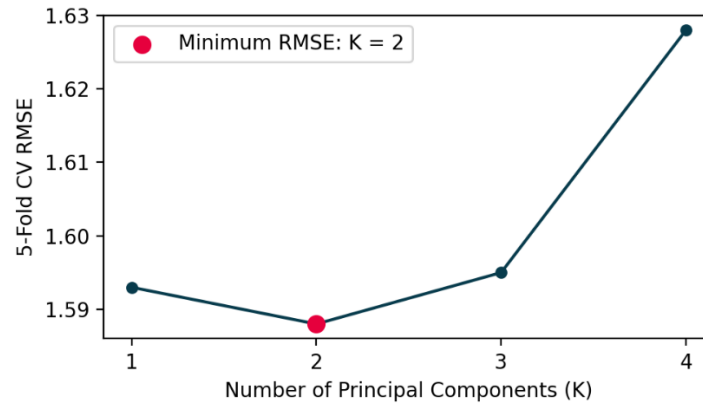


Figure 5. Cross-validation RMSE by the number of principal components.

F. Principal Component Regression Model

The model with two significant components was significant overall ($F = 18.65$; $p = 6.03 \times 10^{-8}$), with an R^2 of 0.2024, adjusted R^2 of 0.1916, RMSE of 1.5435, and MAE of 1.2342 on the modelling data. In Table VII, PC1 was significant (0.6388; $p < 0.001$), whereas PC2 was not individually significant (0.1927; $p = 0.122$); however, we retained PC2 because $K = 2$ produced the best cross-validation RMSE. The cross-validation performance metrics (RMSE 1.5882; MAE 1.2682; R^2 0.1556) were only slightly worse than those for the modelling data, indicating that the model did not suffer from severe overfitting, although its explanatory power was indeed low to moderate.

TABEL VII. KOEFISIEN REGRESI PADA DUA KOMPONEN UTAMA

Parameter	Coefficient	Standard Error	t-Statistic	p-Value
Konstanta	3,1400	0,1273	24,665	< 0,001
PC1	0,6388	0,1081	5,907	< 0,001
PC2	0,1927	0,1240	1,554	0,122

Figure 6 shows that the actual-versus-predicted points are widely dispersed around the diagonal line, consistent with low to moderate explanatory power. The residuals are approximately normal (Jarque–Bera = 0.637; $p = 0.727$), show no autocorrelation (Durbin–Watson = 1.952), and show no systematic pattern in the residuals-versus-predicted plot.

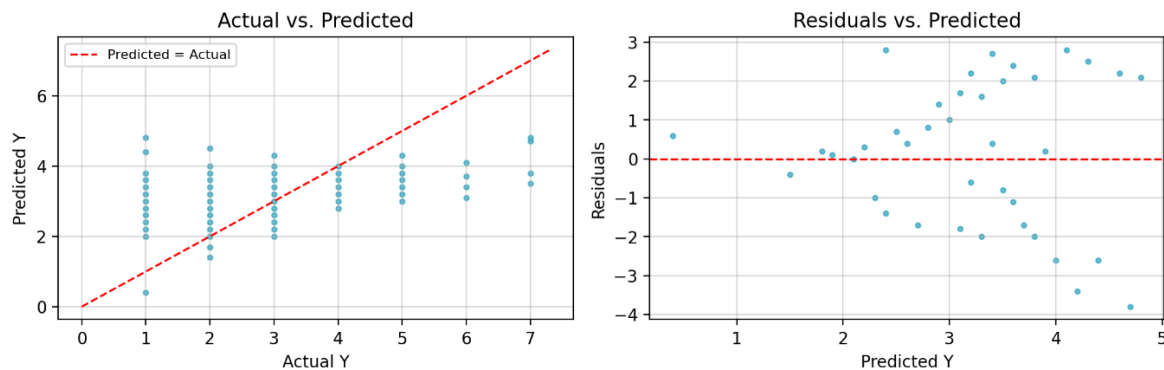


Figure 6. Actual values versus predictions (left) and residuals versus predictions (right).

G. Coefficients on the Original Variable Scale

After transformation to the original scale, Table VIII and Equation (5) show that a one-unit increase in cooking confidence (X1) is associated with an increase in Y of 0.259 times per week, a one-unit increase in X3 with an increase of 0.232, and a one-unit increase in X4 with an increase of 0.150, whereas a one-unit increase in X2 is associated with a decrease in Y of 0.233, with the other variables held constant. The coefficients on the original scale are derived from the component coefficients and therefore do not have their own significance tests. Accordingly, X3 and X1 represent the most consistent associations because they were also significant in the Spearman correlation analysis (Table I), whereas the signs of X2 and X4 should be interpreted with caution because their Spearman correlations were not significant. Figure 7 shows these standardised coefficients: X1, X3, and X4 are positive, and X2 is negative.

TABLE VIII. PCR COEFFICIENTS FOR THE ORIGINAL VARIABLES

Variable	Standardized coefficient	Original scale coefficient	ρ Spearman's rho with Y	p Spearman
X1	0,2713	0,2592	0,2370	0,003
X2	-0,3883	-0,2329	-0,1292	0,115
X3	0,4029	0,2319	0,3987	< 0,001
X4	0,2418	0,1501	0,1354	0,099

$$\hat{Y} = 1,4221 + 0,2592X_1 - 0,2329X_2 + 0,2319X_3 + 0,1501X_4 \quad (5)$$

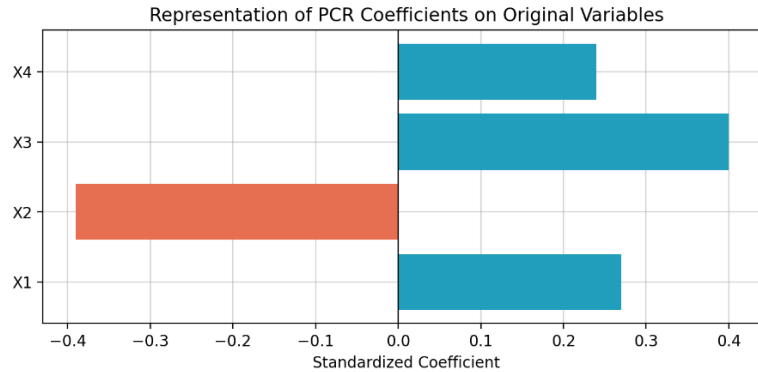


Figure 7. Representation of the standardised PCR coefficients for the original variables.

H. Discussion and limitations

The direction of the associations is consistent with the literature: cooking self-efficacy and cooking frequency increased together among college students who took a food-skills course [5], limited funds were reported as a barrier to eating breakfast [2], and meal-planning behaviour was associated with better diet quality [3]. Thus, the habit of preparing packed meals remains a relevant predictor, although its association in these data was weak. This study has several limitations. Respondents were not selected randomly, and 87.3% came from a single college, so the findings cannot be generalised. All variables were self-reported using discrete scales that were treated as numeric. The researchers specified the four predictors based on their judgment and evaluated them on the same data used for modelling; X2 and X4 were retained even though their Spearman correlations were not significant, so cross-validation performance may be slightly optimistic. We also used cross-validation RMSE to select the number of components and report performance. Finally, the cross-sectional design and low-to-moderate R^2 values mean that the results indicate associations rather than causal relationships.

IV. CONCLUSION

Of the 15 candidate questions, four predictors were selected based on conceptual relevance and supported by significant Spearman correlations for X1 and X3. PCR with two principal components, selected through 5-fold cross-validation and consistent with the Kaiser criterion, explained 61.01% of the predictor variance and produced a simultaneously significant model with an R^2 of 0.2024, an adjusted R^2 of 0.1916, and a cross-validated R^2 of 0.1556 for the frequency of cooking one's own breakfast among 150 college students. On the original scale, cooking confidence, the frequency of cooking because eating out is more expensive, and the frequency of preparing packed meals were positively associated with Y, whereas the frequency of not cooking because ingredients had run out was negatively associated. Cross-validation performance indicated that the model's explanatory power remained low to moderate. Future research should use more diverse, randomly selected samples; add variables such as morning free time, place of residence, and ingredient costs; and establish the predictors before analysing the data.

REFERENCES

- [1] S. Abro, Q. Saleem, J. Lashari, G. Khalid, M. Khan, and F. Malik, "Breakfast practices and factors associated with skipping of breakfast in medical students," *Journal of Rawalpindi Medical College*, vol. 25, no. 1, pp. 37–41, 2021, doi: 10.37939/jrmc.v25i1.1424. [Online]. Available: <https://www.journalrmc.com/index.php/JRMC/article/view/1424>
- [2] R. E. Adonu, M. Amoah, and F. I. Saah, "Breakfast intake and associated factors and barriers among tertiary institution students in the Western Region, Ghana," *BMC Nutrition*, vol. 9, Art. no. 7, 2023, doi: 10.1186/s40795-023-00672-6. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9830603/>
- [3] A. J. Hanson *et al.*, "Cooking and meal planning as predictors of fruit and vegetable intake and BMI in first-year college students," *International Journal of Environmental Research and Public Health*, vol. 16, no. 14, Art. no. 2462, 2019, doi: 10.3390/ijerph16142462. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6679210>

- [4] S. S. Al Jauziyah, Nuryanto, A. F. A. Tsani, and R. Purwanti, "Pengetahuan gizi dan cara mendapatkan makanan berhubungan dengan kebiasaan makan mahasiswa Universitas Diponegoro," *Journal of Nutrition College*, vol. 10, no. 1, pp. 72–81, 2021. [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jnc/article/view/30428>
- [5] C. D. French *et al.*, "Impact of a food skills course with a teaching kitchen on dietary and cooking self-efficacy and behaviours among college students," *Nutrients*, vol. 16, no. 5, Art. no. 585, 2024, doi: 10.3390/nu16050585. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10933768/>
- [6] G. L. Bernardo, M. M. Jomori, A. C. Fernandes, C. F. Colussi, M. D. Condrasky, and R. P. da C. Proença, "Nutrition and Culinary in the Kitchen Program: a randomized controlled intervention to promote cooking skills and healthy eating in university students – study protocol," *Nutrition Journal*, vol. 16, Art. no. 83, 2017, doi: 10.1186/s12937-017-0305-y. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5738807>
- [7] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A*, vol. 374, no. 2065, Art. no. 20150202, 2016, doi: 10.1098/rsta.2015.0202. [Online]. Available: <https://doi.org/10.1098/rsta.2015.0202>
- [8] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021. [Online]. Available: <https://www.statlearning.com>
- [9] M. A. Aprihartha, M. R. Arfani, and I. Putianik, "Implementasi metode regresi dengan analisis komponen utama pada kasus diare balita di Jawa Tengah," *Jurnal EurekaMatika*, vol. 13, no. 2, pp. 123–134, 2025, doi: 10.17509/jem.v13i2.90359. [Online]. Available: <https://ejournal.upi.edu/index.php/JEM/article/view/90359>
- [10] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3, pp. 91–93, 2018, doi: 10.1016/j.tjem.2018.08.001. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6107969>