



UNIVERSITI  
TEKNOLOGI  
MARA

# MATHEMATICS AND STATISTICS

## UNDERGRADUATE RESEARCH PROCEEDINGS 2025

UiTM CAWANGAN NEGERI SEMBILAN



## ANALYZING FOOD SECURITY OF THE HOUSEHOLDS: A CASE STUDY IN MALAYSIA USING A LOGISTIC REGRESSION APPROACH

Julia Johari<sup>1</sup>, Jaida Najihah Jamidin<sup>2\*</sup>

<sup>1,2</sup> Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

\*jaida5698@uitm.edu.my

### Abstract

*Food security is a state where all households have access to sufficient, nutritious, and safe food, physically, socially, and economically. Despite being one of the highest palm oil-producing countries in Asia, Malaysia still shows inconsistent food security. Malaysia's growing population increases food demand but also raises challenges such as food insecurity, fuelled by low income, limited nutritional knowledge, and crisis such as COVID-19. To improve food security among households, the country needs to reduce imports, making food easier to access and affordable. Therefore, this study was conducted to identify the factors that contribute to food insecurity among Malaysians. Logistic regression was handled with independent variables which are gender of household head, household size, the income of the household, access to credit, age of household head, and level of education of household head towards food security. This study was carried out on Malaysian households. The result of this study revealed that only several variables namely household size, income of the household, and level of education contributed significantly to food security in Malaysia.*

**Keywords:** Food Security, Logistic Regression Model

### Introduction

The growing global population increases the demand for various necessities, particularly food. However, due to limited resources and limitless human desire, not every need can be fulfilled. Food plays a vital role in our daily lives, as it is consumed by everyone on a regular basis. It offers a multitude of benefits to consumers, providing essential energy required for various activities and tasks. Without adequate nutrition from food, our bodies would lack the necessary nutrients needed for growth and overall health. Additionally, a well-balanced diet supports proper breathing and enables optimal brain function. In essence, food is indispensable for sustaining life and promoting overall well-being.

Food security is a state in which all households have access to sufficient, nutritious and safe food, physically, socially, and economically. It is also defined as "ensuring that everyone has physical and economic access to the basic food they need at all times" [1]. This is a fundamental necessity that every nation must address to guarantee the access of sufficient food to all individuals. [4] defines household food security as the physical, social, and economic stability that comes from having enough safe and nutritious food available on a consistent basis to meet dietary needs and preferences, thereby promoting a healthy and productive life [4]. Additionally,

food security ensures that food remains affordable despite fluctuations in the country's currency. It is imperative to prioritize food security within households, particularly for individuals with limited incomes who are vulnerable to food insecurity.

In strengthening the country's food security, the National Food Security Framework (KSMN) has developed four main components, namely (1) availability which includes the provision of food stocks, import production, and food aid; (2) access which refers to infrastructure, logistics, and consumers' ability to afford goods; (4) utilization which involves processing and preparation methods that comply with existing standards such as HALAL; and (4) stability which encompasses sustainability and the capacity to withstand crises such as natural disasters, climate change and agricultural diseases ore pests. The main component is in line with the latest concepts recommended by the Food and Agriculture Organization of the United Nations (FAO). The annual rise in population is one of the primary reasons why food insecurity has become a worldwide concern.

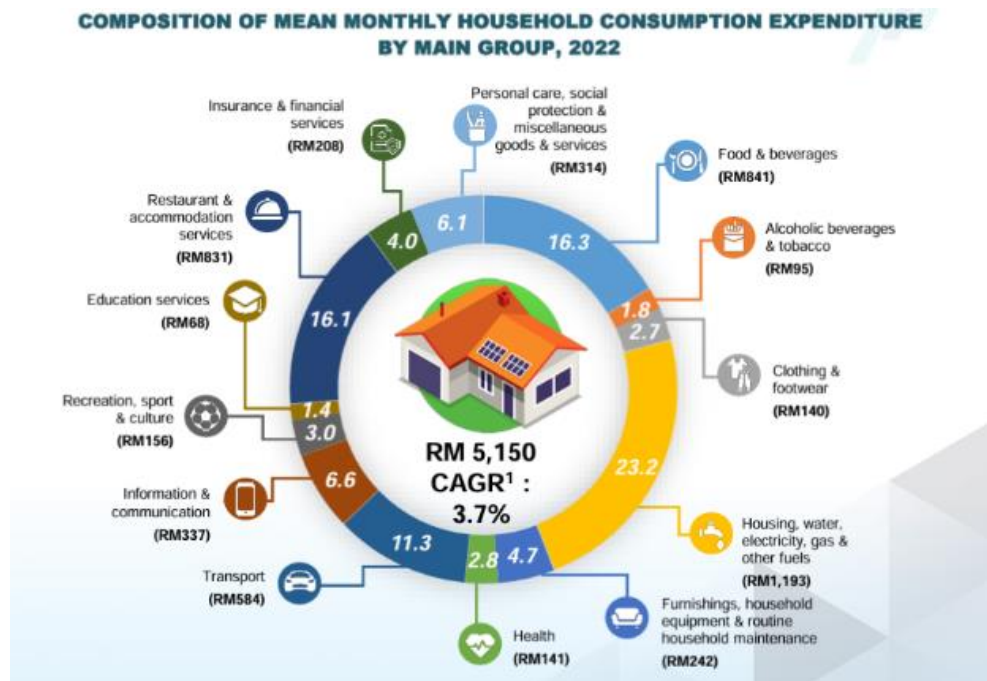


Figure 1: Ranking of Food Security Index based on Countries.

As indicated in Figure 1, Malaysian households spent an average of RM841 per month on food, which represented a significant portion of their expenses [2]. While higher-income families could be secured with this amount, this high burden might influence low-income households in terms of their inability to meet other essential needs. Hence, it pushes them toward food insecurity. These economic challenges could grow further if inflation causes the rise of food prices, particularly among the most vulnerable people in society. As a result, specific measures such as subsidies, food assistance programmes, and price restrictions may be required to ensure that all households can afford and have access to healthy foods, which are the basis of overall health and economic stability.

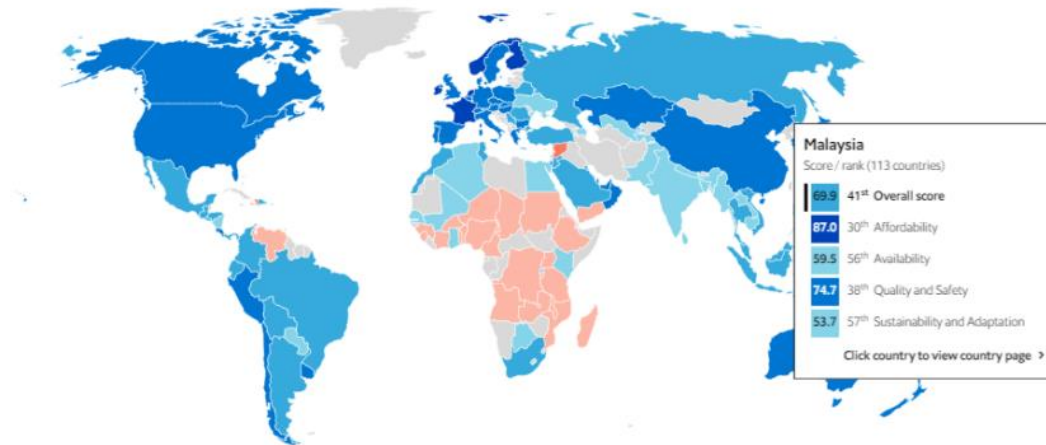


Figure 2: Ranking of Food Security Index based on Countries

Furthermore, based on Figure 2, in 2022 Malaysia is ranked 41 out of 113 countries, which is a drop of seven places compared to the position in 2021 [3]. The situation raises concerns about the security of the availability of food for Malaysian households and how they manage to keep their household food safe. This study therefore seeks to figure out the major factors that contribute to food insecurity.

## Methodology

### *Description of Data*

The data used in this study was collected from a survey of 479 respondents in Malaysia. The dataset includes variables such as household income (continuous, RM), household size (continuous, person), gender of the household head (categorical, male or female), age of household head (continuous, years), and credit card usage (categorical, yes or no). Level of education was coded as "SPM," "Diploma," "Degree," and "Postgraduate". For food security status, 1 if the household food is secure and 0 if the household food is not secure. Besides, the study was conducted from October 2024 to August 2025. The study focused on the Malaysian population aged 18 and above.

### *Logistic Regression Analysis*

Logistic regression analysis was a type of binary classification regression in which the dependent variable had only two possible outcomes. This analysis was used to define variables and determine the relationship between a dependent binary variable and its independent variables. Furthermore, this analysis was distinct from linear regression, which produced a continuous output.



The general form of logistic regression model is:

$$\begin{aligned} \text{logit}(P) &= \log \left[ \frac{P(X)}{1 - P(X)} \right] \\ &= B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K + \varepsilon \end{aligned}$$

The above model can be rearranged to simplify it, and  $P$  can be calculated using the following formula:

$$P = \frac{e(B_0 + B_1X_1 + B_2X_2 + \cdots + B_6X_6 + \varepsilon)}{1 + e(B_0 + B_1X_1 + B_2X_2 + \cdots + B_6X_6 + \varepsilon)}$$

where,

$P$  = Probability of household food is secure

$B_0$  = Constant of the equation

$B_1, \dots, B_k$  = The coefficient of independent variable

$X_1$  = Household Income

$X_2$  = Household Size

$X_3$  = Gender

$X_4$  = Age

$X_5$  = Credit Card Usage

$X_6$  = Level of Education

Five factors were considered for evaluating the model. These criteria were used to assess the logistic regression model. Firstly, Omnibus Test was applied to assess if any regression coefficients significantly differ from all other coefficients. This test is commonly applied in ANOVA and multiple linear regression models. Regarding the hypotheses, if the p-value is less than the alpha value which is set to be 0.05, there was a relationship between the variables. Conversely, if the p-value exceeds 0.05, it indicates there is no relationship between the variables.

Secondly, to test the goodness of fit of a logistic regression model, the Hosmer and Lemeshow test can be used. This statistical test assesses whether the data fits the logit model well. For the regression model to be considered a good fit, the p-value (or significance value) must be greater than 0.05. Conversely, if the p-value is less than 0.05, the regression model is not well fitted. A significant result indicates that the model did not fit the data well, whereas a non-significant result indicates that it does.

Thirdly, predictive efficiency is used for model's ability to accurately predict outcomes or make predictions on new data or unseen data. It measures how well the model performs in terms of predictive accuracy and generalization. A model is considered to have good predictive efficiency if the aggregate percentage was 60% or higher. If it is less than 60%, the model is not considered to have good predictive efficiency. In addition to the overall percentage, sensitivity, specificity, and accuracy can also be used to assess the predictive efficiency of a model. To



calculate these metrics, some terms need to be defined first. True Positive (TP) is when the model accurately predicts the positive class. True Negative (TN) is when the model accurately predicts the negative class. A False Positive (FP) occurred when the model predicts the positive class incorrectly. False Negative (FN) occurs when the model incorrectly predicted a negative class. Sensitivity indicates the model's ability to predict a positive outcome correctly, such as correctly identifying households that were food-secure (coded as '1'). Specificity, on the other hand, measures the model's ability to predict a negative outcome correctly, such as identifying households that are food-insecure (coded as '0'). Accuracy represents the overall correctness of the model and indicates the proportion of correctly classified instances (both positive and negative) out of the total number of instances.

Next, Cox and Snell  $R^2$  and Nagelkerke  $R^2$  were methods that display the amount of variation of the dependent variable that can be explained by independent variables in the current model. These values range from 0 to 1, representing the proportion of variability in the dependent variable that is accounted for by the independent variables.

Lastly, another popular method for testing variable significance in logistic regression is the Wald statistic. It evaluates each regression coefficient separately in order to determine the significance of an independent, regressor, or predictor variable. When the p-value is less than the 0.05, the variable is considered a significant predictor. In contrast, if the p-value exceeds the 0.05, the variable was considered not significant. Thus, significant predictors were included in the regression model, while non-significant predictors are excluded.

## Results and Discussion

### *Criteria for Model Evaluation*

The logistic regression model was evaluated by using five model evaluation including Omnibus Test, Hosmer and Lemeshow Test, Predictive Efficiency Model, and Cox and Snell  $R^2$  and Nagelkerke  $R^2$ .

Table 1: The Results of Criteria for Model Evaluation

<b>The Criteria for Model Evaluation</b>	<b>Full Model</b>
Omnibus Test	0.0000
Hosmer and Lemeshow Test	0.0560
Predictive Efficiency Model (%)	76.10
Cox and Snell $R^2$	0.2720
Nagelkerke $R^2$	0.3750

Based on Table 1, the results highlight the key criteria for the model evaluation. According to the Omnibus Test results, the p-value is 0.000, which is less than the 5% level of significance. This provides sufficient evidence to support the conclusion that improved food security predictions can be made using the information from both independent variables. The Hosmer-



Lemeshow Test results show a p-value of 0.0560, which is greater than the 5% significance level. This suggests that there is insufficient evidence to reject the null hypothesis that the observed and predicted occurrence rates differ significantly. In other words, the model fits the data well. In terms of predictive efficiency, the model achieves an overall predictive accuracy of 76.10%, which exceeds 60%. Therefore, it can be concluded that the model has strong overall predictive efficiency in determining food security status. The results for the Cox and Snell  $R^2$  value are 0.2720, while the Nagelkerke  $R^2$  value is 0.3750. These values indicate that the variation in food security status explained by the model ranges from 27.20% to 37.50%. The model explains food security variations based on factors such as household size, household income, and the education level of the household head (SPM). Additionally, other variables, including gender (female), credit card usage (yes), age of the household head, and higher education levels of the household head (diploma and degree), may further influence the model's predictive value.

### ***Fitting Logistic Regression Model***

The purpose of this study was to determine the best logistic regression model for the food security of households in Malaysia as well as to examine the factors such as gender of household head, household size, household income, credit card usage, age of the household head, level of education (SPM, Diploma, Degree, Postgraduate) that significantly impact this food security status of the households in Malaysia. To simplify and make the conclusion easier to understand, Table 2 was created.

Table 2: Estimated Coefficient and Wald Statistics

<b>Variable</b>	<b>Estimated Coefficient</b>	<b>P-Value for Wald Statistics</b>
Household Income	0.4360	0.0000
Household Size	0.2020	0.0060
Gender of the Household Head	0.3730	0.1200
Age of the Household Head	0.0120	0.2520
Credit Card Usage	0.4640	0.1080
Level of Education (SPM)	-1.1260	0.0210
Level of Education (Diploma)	-0.4300	0.3770
Level of Education (Degree)	-0.6870	0.1670
Constant	-2.0890	0.0670

Table 2 shows that the variables gender of the household head, age of the household head, credit card usage, level of education (Diploma), and level of education (Degree) are five variables that do not statistically significantly explain the food security of the households in Malaysia. This is because, compared to the other variable, their p-value is greater than 0.05. Out of the six factors mentioned in the previous study's literature review, it can be concluded that the three variables household income, household size, and level of education (SPM) are the main factors that



influence food security in households in the respondent's view. The logistic model equation used for the factors that influence food security is as follows.

$$Z = - 2.0890 + 0.4360 \text{ Household Income} + 0.2020 \text{ Household Size} + 0.3730 \text{ Gender of the Household Head(female)} + 0.0120 \text{ Age of the Household Head} + 0.4640 \text{ Credit Card Usage (Yes)} - 1.1260 \text{ Level of Education (SPM)} - 0.4300 \text{ Level of Education (Diploma)} - 0.6870 \text{ Level of Education (Degree)}.$$

## Conclusion

To sum up, the logistic regression analysis was successfully conducted to achieve the objective of identifying the factors which are gender of household head, household size, the income of the household, access to credit, age of household head, and level of education of household head that contributed to food insecurity in the household among respondents. Six assumptions were checked before completing the logistic regression analysis, and all of the assumptions were met. Among all of the factors tested, there were only three significant factors, which were household size, income of the household head, and level of education (SPM).

As a suggestion to future research, identifying additional factors that contribute to household food insecurity in Malaysia remains a valuable area of study. As food security remains a crucial issue nowadays, further study can highlight its significance in ensuring that all citizens have consistent access to adequate, safe, and nutritious food. Furthermore, promoting public awareness through campaigns and educational initiatives can increase the impact of study findings by encouraging shared knowledge and effort to attain national food security.

Conducting long-term analysis is another recommendation for future studies. Weekly, monthly, or yearly data analysis can provide greater insights into food security trends and patterns, allowing households to be classified as food secure or insecure more accurately. Such longitudinal research would provide an understanding of how food security changes in response to diverse socioeconomic and environmental factors. In addition, future research must emphasize the use of larger sample sizes. A larger dataset will produce results that are more relevant and representative of Malaysia's diverse population. Small sample sizes may not capture the full range of factors influencing food security, reducing the robustness of statistical models.

Finally, the adoption of sophisticated robust statistical approaches is strongly recommended. In this investigation, a large number of observations were removed due to the existence of outliers, which influenced the final model. Implementing robust statistical procedures, such as those that are resistant to outliers, may assist in reducing the problem and preserving useful data. For example, techniques such as Mahalanobis distance or other outlier detection methods could be used to successfully identify and manage outliers, ensuring that critical independent variables are kept while logistic regression assumptions remain successfully met. Finally, using probability sampling techniques is advisable to minimize bias.



## References

- [1] Battersby, J. (2012). Beyond the food desert: Finding ways to speak about urban food security in south africa. *Geografiska Annaler: Series B, Human Geography*, 94(2), 141–159.
- [2] Chee, J.-e. (2024, March). What is the average household expenditure in malaysia? [Accessed: 2024-10-08]. <https://dollarsandsense.my/what-is-the-average-household-expenditure-in-malaysia/>
- [3] Global Food Security Index (GFSI). (2022). <https://impact.economist.com/sustainability>
- [4] World Food Summit Declaration. (1996). *Rome declaration on world food security*.