ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 456-466
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Analysis of Taste Preferences Influenced by Environmental Factors using Multinomial Logistics Regression

Muhammad Ihsan Hafizar[1], Norziha Che Him[1]*, Noor Azliza Abd Latif[1], Yusliandy Yusof[2]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hub Pendidikan Tinggi Pagoh KM1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*
[2] *Faculty of Technical and Vocational Education, UTHM Kampus Parit Raja, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA.*

*Corresponding Author: norziha@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.052

Article Info

Received: 4 January 2026
Accepted: 6 June 2026
Available online: 9 July 2026

Keywords

Taste Preferences, Environmental Factors, Chi-square Test, Multinomial Logistics Regression, AIC, BIC, Log-likelihood, Deviance

Abstract

Environmental exposure influences taste preferences through repeated dietary experiences and social learning. Individuals raised in any region with strong culinary traditions often develop preferences aligned with locally available ingredients and traditional cooking taste preferable. For example, communities living in hot and humid climates tend to prefer spicy or strongly flavored foods, partly due to their preservative and appetite-stimulating properties. This study intends to study impact of environmental factors on individual taste preferences, investigating how external surroundings and situational contexts shape culinary choices. This study uses Taste Preferences dataset that dives into a unique synthetic dataset that predicts food taste preferences (Sweet, Spicy, Sour, or Salty) based on lifestyle habits, sleep cycles, exercise levels, climate zones, and cultural cuisine exposure. To analyze relationships, a quantitative approach was employed using a Chi-square test to explore the association between taste preferences and environmental factors, followed by multinomial logistic regression to model the probability of specific taste preferences based on varying environmental predictors. The models predict the probability of a person belonging to a specific "Taste Preference" category compared to a reference category which is the historical cuisine. The model will check using Log-Likelihood, deviance, AIC, BIC. Normally, values of log-likelihood close to zero are indicative of better fit to data Lesser deviance indicates a better-fitting model, AIC and BIC serve as information criteria which researchers use to select their models. Overall, this study provides a scientific bridge between behavioural psychology and sensory science. By using multinomial logistic regression, you are moving beyond just saying "environment matters" to quantifying the probability of how specific surroundings change what we want to eat.

1. Introduction

Human taste preference (TP) is a multidimensional construct influenced by a complex blend of biological, psychological, environmental, and cultural factors. Classical sensory studies have emphasized the role of inherited taste receptor sensitivity that affect the ability to perceive bitterness and sweetness (1). For instance, chronobiology researchers have established that sleep wake cycles can alter hormonal levels such as ghrelin and leptin and affect cravings and flavour sensitivity throughout the day (2).

Environmental exposure, particularly climatic conditions, is another underexplored determinant of taste preference. Studies in ecological psychology and food anthropology suggest that people living in hotter regions tend to prefer spicy and acidic flavours, which may aid in thermoregulation, microbial control, and satiety (3) (4). In contrast, individuals in colder or drier environments often report stronger preferences for salty or umami-rich foods, which may provide both warmth and energy density. These patterns align with the concept of adaptive taste preference the idea that environment shapes not only food availability but also sensory desirability.

In recent years, there has been increasing academic interest in how lifestyle factors and environmental contexts influence human taste preferences. While genetics and cultural exposure remain foundational, emerging research emphasizes that sleep cycles, physical activity levels, and climatic conditions may significantly affect flavour perception and food choices (5) (6). Objective of this study are to explore the association between taste preferences and environmental factors by using chi-square test, develop multinomial logistics regression model for taste and environmental factors, then evaluate the model performance by using AIC, BIC, Log-likelihood, degree of freedom and deviance. The goal of this research to show that taste is not just "in our genes"; it is influenced by where we live and what we do every day. The Usefulness of this information helps people who work in healthcare and food and beverage industry.

2. Method

2.1 Data Description

The first stage of the analysis involved a detailed descriptive investigation of the dataset to characterize the sample population and summarize the distribution of the study variables. The dataset that will be used in this study obtained from an online open data sources called Kaggle website. The research uses secondary data from an open-source dataset titled Flavour Sense. For the continuous variable, Age, standard measures of central tendency (mean and median) and dispersion (standard deviation, minimum, and maximum range) were computed. Categorical variables are Sleep Cycle, Exercise Habits, Climate Zone, Historical Cuisine Exposure, and the dependent variable Preferred Taste were summarized using frequency counts and corresponding percentages. This analysis provided crucial insight into the distribution of the sample across the different categories, noting the high frequencies of Sweet and Sour preferences and the low frequency of Spicy preference in the dependent variables.

2.2 Chi-Square Test

The methodology for employing the Chi-Square test of independence follows a structured, step-by-step statistical framework designed to test hypotheses regarding the association between categorical variables. This process ensures that the transition from raw frequency data to inferential conclusions is mathematically rigorous and theoretically sound.

The initial step involves defining the research parameters through two competing hypotheses. According to (7), the null hypothesis asserts that the two categorical variables are independent (no relationship), while the alternative hypothesis posits that a significant relationship exists between them. Before calculating the test statistic, the researcher must determine the expected frequencies for each cell. This represents the frequency expected if the null hypothesis were true. (8) provides the following formula for calculating the expected frequency for any cell:

Before calculating the test statistic, the researcher must determine the expected frequencies for each cell. This represents the frequency expected if the null hypothesis were true. Agresti provides the Equation (1) to calculate the expected frequency for any cell (8);

$$E_i = \frac{T_t \times T_c}{N} \tag{1}$$

where:

- T_r = Total of the row in which the cell lies.

- T_c = Total of the column in which the cell lies.
- N = Grand total of all observations.

The Chi-Square value is calculated by measuring the squared difference between observed and expected frequencies, normalized by the expected frequency for each cell. According to (7) (9), the formula as in Equation (2);

$$\chi^2 = \sum_{(O_i - E_i)} 2E_i \tag{2}$$

where:

- $\sum$ = The sum across all cells in the table,
- O_i = Observed frequency in cell *i*,
- E_i = Expected frequency in cell *i*.

To determine if the calculated value is statistically significant, the researcher must calculate the degrees of freedom. This is essential for identifying the correct critical value from the Chi-Square distribution table. The resulting p-value is compared against a predetermined significance level (usually $\alpha = 0.05$). If $p < 0.05$, the null hypothesis is rejected. If the test is significant, researchers must calculate the effect size to understand the strength of the association, as the Chi-Square test only indicates that a relationship exists, not how strong it is (10). For tables larger than 2×2, Cramer's V is the standard metric, while the ϕ coefficient is used for 2×2 tables (11) (12).

2.3 Multinomial Logistics Regression

Multinomial Logistic Regression (MLR) was employed to assess how modifiable lifestyle and environmental characteristics influence the likelihood of individuals preferring one of four distinct taste categories. MLR is a generalization of binary logistic regression suitable for modelling nominal outcomes with more than two unordered categories (13).

MLR estimates the log-odds of membership in a category relative to a reference group. The model assumes k categories for the dependent variable. When $k = 1, 2, \ldots, k$, the model estimates use is Equation (3);

$$log\,\frac{Pr(y=k\,|\,x)}{Pr(y=K\,|\,x)} = \beta_0^{(k)} + \beta_1^{(k)}x_1 + \ldots + \beta_p^{(k)}x_p \tag{3}$$

where:

- $P_r(y=k)$: Probability of selecting taste category,
- $\Pr(y=K)$: Probability of selecting the reference category,
- $\beta_p^{(k)}$: Coefficient for predictor   in category.

The model parameters were estimated using Maximum Likelihood Estimation (MLE), a method that determines the coefficient values that maximize the likelihood of the observed data. For the Multinomial Logistic Regression model, the corresponding likelihood function is presented in Equation (4) (14);

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{K} \left[P(Y_i = j)\right]^{y_{ij}} \tag{4}$$

where:

- $y_{ij} = \{ \begin{matrix} 1, if\ observation\ i\ belongs\ to\ category\ j \\ 0, otherwise \end{matrix}$
- $\prod_{j=1}^{K}$ is Product over all possible categories
- $\prod_{i=1}^{n}$ is Product over all observations
- $P(Y_i = j)$ is model-based probability that observation i falls into category j

The regression coefficients were exponentiated to yield odds ratios. An odd ratio greater than 1 indicates that the predictor increases the odds of choosing the specific taste over the reference, while an OR less than 1 indicates a decrease. For example, if the OR for "Night Owl" in the "Sweet" category is 2.5, it implies that Night Owls are 2.5 times more likely to prefer Sweet over Salty, holding other variables constant.

2.4 Model Checking for Multinomial Logistics Regression

The adequacy and stability of the fitted Multinomial Logistic Regression (MLR) model were evaluated through a systematic model checking procedure, which included assessing the overall goodness-of-fit, examining the stability of the estimated coefficients, and identifying potential issues such as multicollinearity and perfect separation.

Likelihood Ratio Test (LRT) was employed to compare the fitted model, which included the predictor variable *Historical Cuisine Exposure*, with a null model containing only the intercept. A statistically significant result ($p < 0.05$) indicates that the fitted model provides a significantly better representation of the data than the null model (7). Information criteria were used to support comparative model evaluation, namely the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria account for model complexity by penalizing the inclusion of additional parameters, with lower AIC and BIC values indicating a more favourable balance between goodness-of-fit and parsimony.

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are statistical metrics used for model selection. They assess a model's goodness of fit while penalising for model complexity to avoid overfitting. The formula for AIC as in Equation (5);

$$AIC = 2k - 2\ln(L) \tag{5}$$

where:

k: The number of estimated parameters in the model (including the intercept)

L: The maximum likelihood of the model (the likelihood of the data given the model parameters)

A lower AIC value indicates a better fit relative to other models.

The formula for BIC as in Equation (6). A lower BIC value indicates a better model fit relative to others. BIC imposes a stricter penalty for complexity than AIC, especially as the number of observations increases (15);

$$BIC = k\ln(n) - 2\ln(L) \tag{6}$$

where:

k: The number of estimated parameters in the model,

n: The number of observations in the dataset,

L: The maximum likelihood of the model.

The unusually large coefficient estimates observed during model development, the data were examined for the presence of quasi-complete separation. Separation occurs when a predictor variable perfectly or near-perfectly predicts the outcome for one or more response categories, resulting in excessively large, unstable, and unreliable coefficient estimates. This issue was diagnosed by inspecting the standard errors (SE) of the estimated coefficients, as abnormally large SE values often in the thousands or tens of thousands are indicative of separation (15).

The accuracy of these classifications was then assessed using standard performance metrics, including a confusion matrix in the form of a 4 x 4 classification table that compared the observed taste categories with those predicted by the model, as well as the overall accuracy, which was calculated as the proportion of correctly classified observations across all taste categories.

3. Results and Discussion

3.1 Descriptive Statistics

Descriptive analysis will show mean, median, max or in value, and standard deviation. Mean is the sum of all values divided by the total count.

Table 1 Descriptive Statistics for Age

Statistic	Age
count	6855
min	18
max	69
median	43

Table 1 provide an overview of the age distribution within a large sample of 6,855 individuals. The study includes a total of 6,855 data points which represent individual people who took part in the research. The youngest person in this dataset is 18 years old. The study investigates adult populations because this requirement exists. The oldest person in the dataset is 69 years old.

Table 2 Frequency and Percentage Distribution of Variables

Variables	Frequency	Percentage
	Sleep Cycle	
Early Bird	2338	34.11
Night Owl	2261	32.98
Irregular	2256	32.91
	Exercise Habits	
Moderate	2340	34.14
Heavy	2267	33.07
Light	2248	32.79
	Climate Zone	
Tropical	1806	26.35
Cold	1724	25.15
Dry	1707	24.9
Temperate	1618	23.6
	Historical Cuisine Exposure	
Mediterranean	2340	34.14
Mixed	2277	33.22
Asian	2238	32.65
	Preferred Taste	
Sweet	3002	43.79
Sour	2515	36.69
Salty	1141	16.64
Spicy	197	2.87

The table demonstrates that participants exhibit equal distribution across their three Sleep Cycles which include Early Bird Night Owl and Irregular Sleep patterns. The study achieves equilibrium between Climate Zones and Cuisine Exposure because Tropical regions and Mediterranean backgrounds represent the two largest demographic groups at 26.35% and 34.14% respectively. People display a strong preference for Sweet which they select at 43.79% of the time while Spicy ranks as the least preferred option chosen by only 2.87% of the people.

3.2 Chi-Square Test

Table 1 shows the Pearson's chi square test for pair of data where it exemplifies a group of statistical association tests aimed at finding out the dependency between human sensory preferences, in particular, Preferred Taste, and a wide array of external and internal variables.

Table 3 Pearson's chi square test for variables pair

Variables pair	x^2	Degree of Freedom	*p*-value
Preferred Taste and Historical Cuisine Exposure	1348.6	6	$p<0.001$
Preferred Taste and Climate Zone	1177.4	9	$p<0.001$
Preferred Taste and Sleep Cycle	1278.5	6	$p<0.001$
Preferred Taste and Exercise Habits	1392.7	6	$p<0.001$

The results of the Pearson's Chi-Square Tests, summarized in Table 1, show that all four variable pairings produced a p-value which reached 2.2^{-16}. In statistical terms, this value is effectively zero, far below the standard threshold of 0.05. The results allow us to dismiss the Null Hypothesis because our confidence level reaches an extremely high point that taste preference depends on these factors. The Chi-Square values show extreme results which range between 1177.4 and 1392.7. The degrees of freedom (6 and 9) reflect the complexity of the categories within each variable. The df values increase to higher levels yet the x^2 statistics maintain their significance which demonstrates that the relationships exist across different sub-groups.

The research results show that an individual's taste preferences depend on three factors which include their lifelong cultural experiences with food, their geographical region's climate, and their bodily natural patterns of sleeping and physical activity. The identical p-value results across different categories demonstrate that taste preference exists as an interdisciplinary trait which people acquire through their genetic heritage, environmental factors, and personal habits.

Table 4 Contingency table for Preferred taste and Historical Cuisine

Preferred taste and Historical Cuisine	Asian	Mediterranean	Mixed	Row Total
Salty	0	587	554	1141
proportion	0	0.251	0.343	
Sour	568	963	984	2515
proportion	0.254	0.451	0.243	
Spicy	563	14	10	197
proportion	0.077	0.006	0.004	
Sweet	1497	776	729	3000
proportion	0.678	0.332	0.322	

Column Total	2248	2347	2277	6855
proportion	0.326	0.341	0.332	1

Table 4 presents the contingency table that shows the distribution of 6855 total respondents across the three exposure categories which are Asian (2238 respondents or 36% of the total), Mediterranean (2340 respondents or 34.1% and Mixed (2277 respondents or 33.2%). The column analysis of the table shows different taste preferences, which reveal that people with Asian exposure show the highest preference for sweet taste (66.9% or 1497 respondents), while both Mediterranean (41.2% or 963 respondents) and Mixed exposure (43.2% or 984 respondents) groups show their largest preference for Sour taste. The Spicy preference shows a higher rate among Asian exposure individuals (7.7% or 173 respondents) than it does for the other two groups where less than 1% of individuals showed that preference.

3.3 Multinomial Logistics Regression

Multinomial logistic regression, a parameter estimate (β) represents the magnitude and direction of the relationship between a predictor variable and the outcome variable. For categorical or binary outcomes, the relationship is expressed in terms of log odds, which is the natural logarithm of the odds of an event occurring versus not occurring.

From value in Tables 5, 6, 7, Multinomial Logistics Regression model builds as shown in Equation (7).

$$Taste(a) vs.Taste(b): \ln\left(\frac{P(.Taste(b))}{P(Taste(b))}\right) = \beta + \beta(\mathrm{HC}\ (\mathrm{a})) - \beta(HC(b)) \tag{7}$$

Table 5 shows the coefficient for MLR with three categories for historical cuisine (HC) as reference category for the independent variable and 4 Taste preferences (TP) as reference category for the dependent variable.

Table 5 Comparative Multinomial Logistic Regression Coefficients by Reference Category (Mixed)

Taste	**coefficient**	**Asian**	**Mediterranean**
Sweet as dependent variables			
Salty	-0.2745	7.8948	-0.0797
Sour	0.2500	9.1641	0.0045
Spicy	-4.2891	11.2897	0.2767
Sour as dependent variables			
Sweet	-0.3000	1.2690	0.3576
Salty	-0.5745	-13.4131	0.0794
Spicy	-4.5887	3.3998	0.0840
Salty as dependent variables			
Sour	0.5746	7.8948	-0.0797
Sweet	0.2749	9.1641	0.0045
Spicy	-4.0094	11.2897	0.2767
Spicy as dependent variables			
Salty	4.0146	-19.0726	-0.2786
Sour	4.5890	-3.4002	-0.3580
Sweet	4.2891	-2.1312	-0.2740

Based on the models, when sweet as the references, the "Mixed" group shows a preference for Sour which exceeds their preference for Sweet by 0.2500. Three people show less preference for Salty and Spicy because their scores show -0.2745 and -4.2891 respectively. The Asian population exhibits a complete shift away from sweet flavours toward all other tastes which includes their preference for Spicy at 11.2897 and Sour at 9.1641.

Mediterranean people prefer Spicy over Sweet by a tiny amount which makes them choose Salty less than the Mixed group does.

When Sour is the reference, we see that the Asian group has a much lower preference for Salty (-13.4131) relative to Sour. The Mixed group shows a preference for Sour above all other categories which is shown through its negative coefficients for Sweet Salty and Spicy. When salty as the references, Mixed group prefers Sour (0.5746) and Sweet (0.2749) over Salty but strongly dislikes Spicy (-4.0094) relative to Salty. For Asian participants, the preference for Spicy over Salty remains extremely high (11.2897), mirroring the trend seen in the sweet baseline section.

When spicy as the references, the Mixed group favours other tastes over Spicy, with high positive intercepts for Salty (4.0146), Sour (4.5890), and Sweet (4.2891). The Asian group shows a complete reversal because their negative coefficients show that they prefer Spicy more than any other group when compared to these alternatives. Overall, for the Mixed group, Sour taste was the most preferred choice in all model tests that used Sweet or Salty or Spicy as their baseline. The participants demonstrated mild preferences for Sweet and Salty, but they maintained a strong dislike for all Spicy flavours

Table 6 Comparative Multinomial Logistic Regression Coefficients by Reference Category (Asian)

Taste	coefficient	Mixed	Mediterranean
	Sweet as dependent variables		
Salty	-14.1552	13.8806	13.8761
Sour	-0.9691	1.269	1.185
Spicy	-2.1578	-2.1311	-1.8575
	Sour as dependent variables		
Sweet	0.9691	-1.2691	-1.185
Salty	-13.2785	12.704	12.7834
Spicy	-1.1888	-3.4002	-3.0422
	Salty as dependent variables		
Sour	13.0968	-12.5223	-12.6017
Sweet	14.0659	-13.7914	-13.7867
Spicy	11.908	-15.9227	-15.6438
	Spicy as dependent variables		
Salty	-13.5746	17.5892	17.3106
Sour	1.1888	3.4002	3.0422
Sweet	2.1579	2.1312	1.8572

Based on models for Asian as the references, when sweet as the references, Asian group members show least favourite taste for all flavours except Sweet that they prefer above all other flavours. The Mixed and Mediterranean groups show strong preference for Salty taste because their members have become over twice as likely to choose Salty taste instead of sweet taste when compared to Asian group members.

When sour as the references, The Asian group shows a slight preference for Sweet over Sour (0.9691) but a very strong dislike for Salty (-13.2785). The Mixed and Mediterranean groups demonstrate strong preference for Salty when compared to Sour with coefficients approximately 12.7 while the Asian group shows different results. When salty as references, the three other flavours which include Sour and Sweet and Spicy flavours all receive greater preference than Salty flavouring. Sweet flavouring received the highest positive value at 14.0659, which represents its most favourable taste assessment. The Mixed and Mediterranean groups show extreme disfavour towards all flavours except Salty because their negative coefficients reach between -12.5 and -15.9.

When spicy as the references, The Asian group prefers Sour (1.1888) and Sweet (2.1579) over Spicy, but they have a strong dislike for Salty (-13.5746). The Mixed and Mediterranean groups display their most intense taste for Salty (17.5892 and 17.3106) while showing they value Sour and Sweet at lower levels when compared to Asian baseline. Overall, for the Asian group consistently ranks Sweet as their most favorable taste. The preference shows statistically dominant results through testing both Sweet and other flavours. They show strong and ongoing dislike for Salty tastes which people consider less appealing than all other taste types.

Table 7 Comparative Multinomial Logistic Regression Coefficients by Reference Category (Mediterranean)

Taste	coefficient	Mixed	Asian
	Sweet as dependent variables		
Spicy	-4.0148	1.8569	-0.2745
Salty	-0.2791	-14.3918	0.0046
Sour	0.2159	-1.185	0.084
	Sour as dependent variables		
Sweet	-0.2159	1.185	-0.084
Spicy	-4.2307	3.0419	-0.3583
Salty	-0.495	-13.59	-0.0794
	Salty as dependent variables		
Sour	0.495	8.4801	0.0794
Sweet	0.2791	9.6651	-0.0046
Spicy	-3.7328	11.5192	-0.2774
	Spicy as dependent variables		
Salty	3.736	-18.6738	0.2786
Sour	4.231	-3.0422	0.358
Sweet	4.0151	-1.8572	0.274

Based on models for Mediterranean as the references, Spicy and Salty are less preferred than Sweet (coefficients of -4.0148 and -0.2791), while Sour is preferred slightly more than the other two options. The Mixed group shows a stronger tendency to select Spicy over Sweet when compared to the Mediterranean group because their odds ratio stands at 1.8569 while their preference for Salty decreases by -14.3918. The Asian group shows very little difference from the Mediterranean group in these specific log-odds.

When sour as the references, the Mediterranean group shows a general dislike for other tastes compared to Sour, which they find more appealing than Spicy (-4.2307). The Mixed group shows a stronger preference for Sweet (1.185) and Spicy (3.0419) than the Mediterranean group, but they show a massive aversion to Salty (-13.59). When salty as the references, the Mediterranean baseline shows that people prefer Sour (0.495) and Sweet (0.2791) but they strongly dislike Spicy (-3.7328). The Mixed group shows a higher preference for all three other tastes (Sour, Sweet, and Spicy) than the Mediterranean group shows for Salty, with Spicy showing the largest increase (11.5192).

When spicy as the references, all other flavours are preferred over Spicy, particularly Sour (4.231) and Sweet (4.0151). The Mixed group, however, is much less likely than the Mediterranean group to choose Salty (-18.6738), Sour (-3.0422), or Sweet (-1.8572) when the alternative is Spicy. Overall, the Mediterranean group shows their highest taste preference for Sour across all models which they follow with Sweet as their second most preferred taste. When Sweet is the baseline, Sour is the only flavour with a positive coefficient, while When Sour is the baseline, all other flavours (Sweet, Salty, Spicy) have negative coefficients, confirming Sour is the "peak" preference. The Mediterranean group shows a statistically significant "strong dislike" for Spicy flavours. The baseline does not affect Spicy, which always receives its worst ratings through negative coefficients.

3.4 Model Checking

The model checking phase focuses exclusively on the sweet taste category and the Mixed cuisine exposure group. This selective verification proves that the mathematical optimization process, which successfully reached convergence, remains stable and reliable for the core predictors of the study. Table 8 presents the model checking metric that use to evaluate the model:

Table 8 Model Checking Metric

Dependent Variable	Log-Likelihood	df	deviance	AIC	BIC	Accuracy
			Mixed as the references			
Spicy	-2.211302698	33	4.422605396	70.4226054	295.9028139	1
Salty	-0.13253717	33	0.26507435	66.2650743	291.745283	1
Sour	-1.53629006	33	3.07258011	69.0725801	294.552789	1
Sweet	-0.89527045	33	1.79054091	67.7905409	293.270749	1
			Asian as the references			
Spicy Model	-2.02034908	33	4.04069816	70.0406982	295.520907	1
Salty Model	-1.1583801	33	2.3167602	68.3167602	293.796969	1
Sour Model	-1.31510636	33	2.63021273	68.6302127	294.110421	1
Sweet Model	-1.18589906	33	2.37179812	68.3717981	293.852007	1
			Mediterranean as the references			
Spicy Model	-2.67725068	33	5.35450135	71.3545014	296.83471	1
Salty Model	-7.6369E-05	33	0.00015274	66.0001527	291.480361	1
Sour Model	-0.00026381	33	0.00052762	66.0005276	291.480736	1
Sweet Model	-0.46441978	33	0.92883956	66.9288396	292.409048	1

Table 8 presents a model comparison and goodness-of-fit analysis for several logistic or multinomial regression models, which the researchers divided into three reference groups: Mixed Asian and Mediterranean.

The measures of log-likelihood capture the fit found in this case; normally, values close to zero are indicative of better fit to data. There, the degrees of freedom were at a constant 33 across all models, suggesting the complexity (number of parameters) remains the same. Within the area of statistics, deviance is an instructive measure, where lesser deviance indicates a better-fitting model. AIC and BIC serve as information criteria which researchers use to select their models. The preferable outcome for the evaluation process requires assessment results to show lower values. The testing system applies penalties to models which include excessive variables because this practice leads to development of overfit models.

Based on the table, The Salty model provides the best fit because it achieved the lowest AIC score of 66.27 and the lowest Deviance value of 0.265 when we used Mixed as the reference. The Spicy model performs the worst in this group (AIC of 70.42). Asian as the Reference, the models in this group display consistent behaviour because their AIC values maintain stability between 68 and 70. Salty and Sweet models show better performance than Spicy and Sour models for this situation. While when Mediterranean as the references category, the Salty and Sour models show their strongest performance because they achieve their highest Log-Likelihood values which approach zero and their lowest AIC values which equal 66.00. The Salty model here has a Deviance of only 0.00015 which shows its data point matching accuracy to an almost perfect level.

The Salty Model and Sour Model (under the Mediterranean reference) show better statistical performance because their AIC values reach a minimum point of 66.00. The two models show the best performance because their AIC values reach the lowest point of the table. The Salty Model shows a deviance value of 0.00015 while the Sour Model shows a deviance value of 0.00052. The models show almost complete accuracy in their predictions because their values approach zero. The highest log-likelihood values for both models show their statistical accuracy reaches near perfection.

The Spicy models across all reference groups are generally the weakest. The models always maintain their highest AIC and BIC values which results in an AIC level of 71.35 for the Mediterranean group. The AIC value shows that the "Spicy" data patterning becomes more difficult for the model to explain than any other flavour which the model handles. The system shows less efficiency when it attempts to make predictions.

4. Conclusion and Recommendations

Based on all result from the analysis, for the Chi-Square values, it shows extreme results, while the df values increase to higher levels yet the x^2 statistics maintain their significance which demonstrates that the relationships exist across different sub-groups. The Mixed group showed Sour taste as their top preference through all model tests which used Sweet or Salty or Spicy as their baseline. The participants showed slight preference for Sweet and Salty, yet they maintained complete dislike for all Spiciness levels. The Asian group shows Sweet as their top flavour preference throughout all testing procedures. The preference shows statistically dominant results through testing both Sweet and other flavours. The Mediterranean participants display a strong ongoing dislike for Salty tastes which they find less appealing than all other taste types. The Mediterranean group displays their strongest taste preference for Sour taste through all models while sweet taste stands as their second most preferred flavour. The study shows a statistically significant "strong dislike" for Spicy flavours according to its findings. The model results show Salty Model and Sour Model (under the Mediterranean reference) demonstrate better statistical performance than Spicy models which exist across all reference groups for model checking. In conclusion this research demonstrates that flavour preferences show strong differences between various cultural groups. The Asian group exhibits a "Sweet-first" flavour profile while the Mediterranean and Mixed groups show a preference for "Sour" flavours and an aversion to "Salty" tastes. The study results receive validation through multiple statistical models which demonstrate better predictive performance through Salty and Sour models than Spicy models. The data shows that sweetness maintains widespread attraction while sourness and spiciness taste preferences show strong division among different subgroups.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, for its support.

Conflict of Interest

The authors declare no conflict of interest in the paper's publication

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Muhammad Ihsan;* ***solve the governing equation:*** *Muhammad Ihsan, Norziha Che Him;* ***data collection:*** *Muhammad Ihsan;* ***analysis and interpretation of results:*** *Muhammad Ihsan, Norziha Che Him;* ***draft manuscript preparation:*** *Muhammad Ihsan, Norziha Che Him, Noor Azliza Abd Latif, Yusliandy Yusof. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Rozin, P. (2006). The integration of biological, social, cultural and psychological influences on food choice. In *The psychology of food choice* (pp. 19–39). CABI. https://doi.org/10.1079/9780851990323.001

[2] McHill, A. W., & Wright, K. P. (2017). Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. Obesity Reviews, 18(S1), 15–24. https://doi.org/10.1111/obr.12503

[3] Billing, J., & Sherman, P. W. (1998). Antimicrobial Functions of Spices: Why Some Like it Hot. The Quarterly Review of Biology, 73(1), 3–49. https://doi.org/10.1086/420058

[4] Prescott, J. (2015). Multisensory processes in flavour perception and their influence on food choice. Current Opinion in Food Science, 3, 47–52. https://doi.org/10.1016/j.cofs.2015.02.007

[5] McHill, A. W., & Wright, K. P. (2017). Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. Obesity Reviews, 18(S1), 15–24. https://doi.org/10.1111/obr.12503

[6] Spence, C. (2015). Multisensory Flavor Perception. Cell, 161(1), 24–35. https://doi.org/10.1016/j.cell.2015.03.007

[7] van der Laan, L. N., de Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. NeuroImage, 55(1), 296–303. https://doi.org/10.1016/j.neuroimage.2010.11.055

[8] Field, A. (2018). Discovering statistics using IBM SPSS statistics ((5th ed.)). SAGE Publications

[9] Agresti, A. (2018). An introduction to categorical data analysis ((3rd ed.)). Wiley.

[10] McHugh, M. L. (2013). The Chi-square test of independence. Biochemia Medica, 143–149. https://doi.org/10.11613/BM.2013.018

[11] Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152. https://doi.org/10.5395/rde.2017.42.2.152

[12] Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

[13] Sharpe, D. (2015). Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research, and Evaluation*, *20*(1), 8.

[14] Hosmer, D. W., Lemeshow, Stanley., & Sturdivant, R. X. . (2013). Applied logistic regression. Wiley.

[15] Greene, W. H. (2003). *Econometric analysis*. Prentice Hall.

[16] Long, J. S. (1997). Regression models for categorical and limited dependent variables. SAGE Publications