Applied Mathematics and Computational Intelligence
Volume 14, No. 4, 2025 [133 - 143]

Multinomial Logistic Regression Analysis of Smoking Status and Demographic Factors in Malaysia

Khatijahhusna Abd Rani[1*], Hamzah Abdul Hamid[2], Maz Jamilah Masnan[3], Wan Zuki Azman Wan Muhamad[4]

[1,2,3,4]Department of Mathematical Sciences, Faculty of Intelligent Computing, Universiti Malaysia Perlis, Pauh Putra Main Campus, 02600 Arau, Perlis, Malaysia

[*] Corresponding author: khatijahhusna@unimap.edu.my

Received: 6 September 2024
Revised: 18 October 2024
Accepted: 24 September 2025

ABSTRACT

This study investigates the association between demographic factors and smoking behavior among individuals in Malaysia, utilizing a secondary dataset of 643 subjects categorized as daily smokers, ex-smokers, and non-smokers. The analysis focuses on eight independent variables: weight, height, gender, ethnicity, marital status, education level, occupation, and physical activity. Descriptive statistics are employed to summarize the characteristics of the sample, while multinomial logistic regression is employed to identify significant predictors of smoking status. The results indicate that the final model, which includes predictors, significantly improves the fit compared to the intercept-only model (p-value < 0.05). This study identified height, gender, ethnicity, and education level as significant predictors of daily smoking behavior. Males and individuals from the Malay ethnic group show higher odds of being daily smokers, while higher education was associated with lower smoking rates. However, the goodness of fit test (p-value <0.001), indicates that the model does not fit the data well. This difference suggests that while the model captures some important relationships among variables, it fails to function as a predictive model for smoking behavior.

Keywords: Multinomial logistic regression, Smoking Status, Demographic Factors.

1 INTRODUCTION

Smoking is one of the major reasons for diseases and death that are classified as preventable all over the world. Understanding how smoking status varies across different demographic groups can reveal behavioral patterns within specific ethnic and cultural communities and guide future strategies [1], [2]. Recent studies have explored the impact of various factors such as gender, educational level, socioeconomic status, physical activity, weight, height, and gender separately on smoking behavior [1]-[12]. For example, individuals with a lower educational status are often at a higher risk of smoking and are less likely to quit. While individuals with smoking status is often associated with lower physical activity levels and unhealthy dietary habits. Thus, analyzing smoking status in relation to these factors, can help researchers to identify the population at risk that may require additional support and resources. This paper aims to identify all factors influencing smoking behavior among people in Malaysia by applying multinomial logistic regression as a statistical method.

2 MATERIAL AND METHODS

2.1 Dataset

In this study, we utilized a secondary dataset comprising 643 subjects. The dependent variable has three categories: daily smoker, ex-smoker and non-smoker. The independent variables selected for this analysis were chosen based on their potential influence on smoking behavior. There are eight independent variables: weight; height; gender (male, female); ethnicity (Malay, others), marital status (married, not married); education level (tertiary, secondary, primary); occupation (private employee, government/ semi government, self-employed); and physical activity (active, inactive).

2.2 Statistical Analysis

Descriptive statistical analyses were employed to determine the basic characteristics of the subjects, providing simple summaries of the sample and measurements [13]. Continuous variables (weight and height) were described by comparing their means, medians, interquartile ranges (IQRs), and standard deviations across three levels of smoking status (Daily smoker, Ex-smoker, and Non-smoker). Categorical variables were summarized using counts and percentages. Then, multinomial logistic regression analysis was employed to fit the model and test the association between smoking status and the identified demographic factors.

2.3 Multinomial Logistic Regression

A regression model is a statistical technique used to examine the relationship between one or more independent variables and a dependent variable. The goal is to make predictions or inferences about the dependent variable based on the values of the independent variables. When the dependent variable (Y) is categorical, a logistic regression model is appropriate.

The choice of the logistic regression model depends on the specific characteristics of the categorical variable. The most common types of logistic regression include Binary Logistic Regression, Multinomial Logistic Regression and Ordinal Logistic Regression. When the dependent variable is binary, meaning it has only two possible categories (e.g., yes/no, success/failure), binary logistic regression is employed. Multinomial logistic regression is a type of logistic regression that is used when the outcome variable has more than two categories that are not ordered. When the outcome variable has more than two categories and these categories have a natural order or classification (e.g., low, medium, high), ordinal logistic regression can be utilized.

This study considers smoking status as a dependent variable with three categories. Multinomial logistic regression is employed to fit the model. The following steps were undertaken:

a) Identify Variables: The multinomial logistic regression model was specified with smoking status as the dependent variable and demographic factors as independent variables. The reference category for smoking status was set as non-smokers.

b) Model Fitting: The model was fitted using the maximum likelihood estimation method. After fitting the model, the model fit was assessed using pseudo-R-squared, and the goodness-of-fit test proposed by[14], which is based on a clustering partitioning strategy. The proposed goodness-of-fit test is suitable because the data includes continuous variables. The analyses were conducted using SPSS software and R, an open-source programming language (version 4.4.1).

c) Interpretation of results: The odds ratio (OR) for each significant demographic factor was calculated, and statistical significance was determined if the p-value was less than 0.05. The OR is used to interpret the impact of each predictor variable on the probability of being in a specific outcome category compared to a reference category.

Multinomial logistic regression extends the concept of binary logistic regression to accommodate dependent variables with more than two categories. While binary logistic regression is designed to predict outcomes with two possible values, multinomial logistic regression allows for the analysis of scenarios where the outcome can fall into multiple, unordered categories. Despite this extension, both models share a similar underlying structure and assumptions, such as the use of maximum likelihood estimation and the modeling of log-odds as a linear function of the independent variables [15]. The following sections present the models of logistic regression.

The single logistic regression model with one independent variable X is given by [16]:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{1}$$

where:

$\pi(x)$ represents the conditional mean of Y given x when the logistic regression is used.

β_0 is the intercept

β_1 is the coefficient for the independent variable x

A transformation of $\pi(x)$ that is central to the study of logistic regression is the logit transformation.

This transformation is defined, in terms $\pi(x)$, as [16]

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{2}$$

This model estimates the probability of a binary outcome based on one predictor.

The multinomial logistic regression model is an extension of binary logistic regression that can handle dependent variables with more than two categories. Let's assume the dependent variable Y has *J* categories.

The model is defined as:

$$\ln\left(\frac{\pi(x)_j}{\pi(x)_J}\right) = \beta_0 + \beta_{j1}x_1 + \beta_{j2}x_2 + \ldots + \beta_{jk}x_k \text{ for } j = 1, 2, \ldots, J-1, \quad (3)$$

where:

$\pi(x)_j$ is the probability of Y being in category j

$\pi(x)_J$ is the probability of Y being in the reference category J

β_0 is the intercept for category j

β_{jk} is the coefficient for the j^{th} logit and the k^{th} independent variable

This model estimates the log-odds of being in each category relative to the reference category as a linear function of the independent variables. The connection between multinomial logistic regression and binary logistic regression is that when $J = 2$ (two categories), the multinomial logistic regression reduces to the binary regression model. In this case, there is only one logit, and the model simplifies to:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \ldots + \beta_{jk}x_k \quad (4)$$

Which is the same with multinomial logistic regression model. Therefore, multinomial logistic regression is a generalization of binary logistic regression that can handle multiple categories in the dependent variable.

Thus, the conditional probabilities of each outcome category in this study can be expressed as follows by referring [16]-[17]:

For Smoking Status 0 ("Daily Smoker"):

$$P(Y=0 \mid x) = \frac{e^{\beta_{00}+\beta_{01}x_1+\beta_{02}x_2+\ldots+\beta_{08}x_8}}{1+e^{\beta_{00}+\beta_{01}x_1+\beta_{02}x_2+\ldots+\beta_{08}x_8}+e^{\beta_{10}+\beta_{11}x_1+\beta_{12}x_2+\ldots+\beta_{18}x_8}} \quad (5)$$

For Smoking Status 1 ("Ex-Smoker"):

$$P(Y=1\Big|x)=\frac{e^{\beta_{10}+\beta_{11}x_1+\beta_{12}x_2+\ldots+\beta_{18}x_8}}{1+e^{\beta_{00}+\beta_{01}x_1+\beta_{02}x_2+\ldots+\beta_{08}x_8}+e^{\beta_{10}+\beta_{11}x_1+\beta_{12}x_2+\ldots+\beta_{18}x_8}} \quad (6)$$

For Smoking Status 2 ("Non-Smoker") as reference category:

$$P(Y=2\Big|x)=\frac{1}{1+e^{\beta_{00}+\beta_{01}x_1+\beta_{02}x_2+\ldots+\beta_{08}x_8}+e^{\beta_{10}+\beta_{11}x_1+\beta_{12}x_2+\ldots+\beta_{18}x_8}} \quad (7)$$

The model specifies the log-odds for the other categories (daily smoker and ex-smoker) relative to the reference category (non-smoker). The probabilities for each category also can be derived from these log odds. The coefficients obtained by fitting the model can then be interpreted to understand the relationship between the independent variables and the likelihood of each outcome.

3 RESULTS AND DISCUSSION

3.1 Descriptive Analysis

The dataset contains eight independent variables, which include both categorical and continuous variables. Table 1 presents all the categorical variables in the dataset, while Table 2 and Table 3 display the two continuous independent variables. Additionally, Table 2 and Table 3 illustrate how descriptive statistics for the continuous variables are typically presented.

Table 1 summarized the frequency distribution of independent variables across the dependent variables. The Cross Tabulation [18] is used to display the information.

Table 1 : Frequency Distribution of the Dataset

Variable Name	Variable Description	Frequency (Percentage)		
		Daily Smoker	Ex-Smoker	Non-Smoker
gender	male (0)	186 (46.6)	23 (5.8)	190 (47.6)
	female (1)	2 (0.8)	3 (1.2)	239 (98.0)
ethnicity	Malay (0)	168 (30.5)	23 (4.2)	359 (65.3)
	others (1)	20 (21.5)	3 (3.2)	70 (75.3)
marital status	married (0)	147 (29.8)	21 (4.3)	325 (65.9)
	not married (1)	41 (27.3)	5 (3.3)	104 (69.3)
education level	tertiary (0)	9 (7)	3 (2.3)	117 (90.7)
	secondary (1)	118 (32.5)	16 (4.4)	229 (63.1)
	primary (2)	61 (40.4)	7 (4.6)	83 (55.0)
occupation	private employee (0)	49 (28.8)	4 (2.4)	117 (68.8)
	government/ semi government (1)	20 (14.9)	5 (3.7)	109 (81.3)
	self-employed (2)	119 (36.2)	7 (2.1)	203 (61.7)
physical activity	active (0)	171 (31.5)	18 (3.3)	354 (65.2)
	inactive (1)	17 (17.0)	8 (8.0)	75 (75.0)

Among the subject, the majority are male, with 186 daily smokers (46.6%), 23 ex-smokers (5.8%), and 190 non -smokers (47.6%). The ethnic composition shows that Malay is dominant with 168 daily smokers (30.5%), while 23 ex-smokers (4.2%) and 359 non-smokers (65.3%). In terms of marital status, 147 daily smokers (29.8%) are married, while 21 ex-smokers (4.3%) and 325 non-smokers (65.9%) also fall into this category. Regarding education, most smokers have secondary education level. Non-smokers show a high level of tertiary education as well (117 or 90.7%). Occupational status reveals that out of 188 daily smokers, 119 are self-employed, 49 are private employees, while 20 work in government or semi-government roles. Physical activity levels show that 171 daily smokers (31.5%) are classified as active, while 18 ex-smokers (3.3%) and 354 non-smokers (65.2%) are also active. Overall, the demographic analysis based on the table above indicates the differences in gender, ethnicity, marital status, education level, occupation, and physical activity among daily smokers, ex-smokers, and non-smokers. These findings suggest that smoking behaviors may be influenced by a combination of social, economic, and demographic factors.

Table 2 and Table 3 present the descriptive statistics for weight and height across the three smoking-status groups: daily smokers, ex-smokers, and non-smokers. Ex-smokers recorded the highest mean weight (75.03 kg), followed by daily smokers (69.56 kg) and non-smokers (68.08 kg)

The minimum weights range from 34.10 kg for non-smokers to 42.40 kg for ex-smokers, while the maximum weights range from 107.30 kg for ex-smokers to 132 kg for daily smokers. The standard deviation of weight is slightly similar across the groups, ranging from 15.41 kg for non-smokers to 16.66 kg for daily smokers. Daily smokers have the highest average height at 166.83 cm, followed closely by ex-smokers at 166.12 cm, while non-smokers have the lowest average height at 159.92 cm. Based on descriptive statistics below, suggest that ex-smokers tend to have the highest average weight, while non-smokers have the lowest average weight and height. Daily smokers are taller on average than non-smokers but slightly shorter than ex-smokers.

Table 2 : Descriptive Statistics (Weight)

Variable	Mean	Minimum	Maximum	Std. Deviation
Daily Smoker	69.56	38	132	16.66
Ex-Smoker	75.03	42.4	107.30	15.62
Non-Smoker	68.08	34.10	128.50	15.41

Table 3 : Descriptive Statistics (Height)

Variable	Mean	Minimum	Maximum	Std. Deviation
Daily Smoker	166.83	149.5	196	7.17
Ex-Smoker	166.12	147.50	192	9.48
Non-Smoker	159.92	135.40	184.60	8.32

3.2 Multinomial Logistic Regression

Table 4 shows the model fitting information for Multinomial Logistic Regression. The results indicate that the final model, which includes predictors, significantly improves the fit compared to the intercept-only model. The Chi-Square test result of 286.382 with a p-value <0.05 confirms that the model is statistically significant, indicating that at least one significant predictor in the model.

Table 4 : Model Fitting Information

	Model Fitting Criteria		**Likelihood Ratio Tests**	
Model	-2 Log Likelihood	Chi-Square	df	*p-value*
Intercept Only	976.409			
Final	690.027	286.382	20	<0.0001

Table 5 presents the pseudo-R-squared values. The Nagelkerke pseudo-R-squared is used in this study due to its ability to scale the value to encompass the entire range of 0 to 1, which facilitates its interpretation [19]-[20]. This is a modified version of the Cox and Snell R-squared. Values that are closer to 1 indicate a more satisfactory fit, similar to R-squared in linear regression. Based on the result, a Nagelkerke pseudo-R-squared value of 0.46 suggests that the model explains approximately 46% of the variability in the outcome variable, relative to a null model (a model with no predictors). Table 6 shows the result of the goodness-of-fit test proposed by [14]. The p-value of <0.001 indicates that the model does not fit the data well.

Table 5 : Model Fitting Information

Pseudo R-squared	**value**
Nagelkerke	0.460

Table 6 : Model Fitting Information

Test	**Value**
χ^2_{p*10}	3517.74

Table 7 summarizes the multinomial logistic regression results. Based on Table 7, the estimated models are written as follows:

For Smoking Status 0 ("Daily Smoker"):

$$\ln\left[\frac{P(Y=0|x)}{P(Y=2|x)}\right]$$
$$= -11.04 + 0.04[\text{Height}] + 4.30[\text{Gender} = \text{Male}] + 1.08[\text{Ethnicity} = \text{Malay}] - 2.19[\text{Education} = \text{Tertiary}]$$

For Smoking Status 1 (“Ex-Smoker”):

$$\ln\left[\frac{P(Y=1|x)}{P(Y=2|x)}\right]$$
$$= -9.59 + 1.88[\text{Gender} = \text{Male}] - 1.10[\text{Physical activity} = \text{Active}]$$

Based on the results, height, gender, ethnicity, and education level are significant predictors of daily smoking behavior. The odds ratio for height (OR = 1.04) indicates that taller individuals are more likely to smoke compared to shorter individuals. Being male (coded as 0) significantly increases the odds of being a daily smoker by approximately 73.8 times compared to females (coded as 1). This finding highlights a strong gender difference in smoking behavior, with males being much more likely to smoke daily.

Individuals from the Malay ethnic group (coded as 0) have nearly 2.95 times higher odds of being daily smokers compared to the other ethnic group. This suggests that ethnicity plays a significant role in smoking behavior, with certain ethnicities being more likely to smoke.

Higher education levels (coded as 0) are associated with a significant decrease in the odds of being a daily smoker, with an odds ratio of 0.11. This finding highlights the importance of education in smoking prevention, suggesting that individuals with higher educational levels are much less likely to smoke. For ex-smokers, gender and physical activity were identified as significant predictors. The odds ratio for gender (OR = 6.56) indicates that males are approximately six times more likely to be ex-smoker than females. Furthermore, individuals who are physically active are less likely to be ex-smokers compared to non-smokers.

Table 7 : Model Fitting Information

Smoking Status	Variable	B	Exp(B)
Daily Smoker	Intercept	-11.04	
	Weight	-0.01	0.99
	Height	0.04*	1.04
	[Gender=.00]	4.30*	73.79
	[Gender=1.00]	0b	.
	[Ethnicity=.00]	1.08*	2.95
	[Ethnicity=1.00]	0b	.
	[Marital_status=.00]	-0.16	0.85
	[Marital_status=1.00]	0b	.
	[Education=.00]	-2.19*	0.11
	[Education=1.00]	-0.51	0.60
	[Education=2.00]	0b	.
	[Occupation=.00]	0.09	1.09
	[Occupation=1.00]	-0.33	0.72
	[Occupation=2.00]	0b	.
	[Physical_activity=.00]	0.33	1.39
	[Physical_activity=1.00]	0b	.

Ex-Smoker	Intercept	-9.59	
	Weight	0.01	1.01
	Height	0.04	1.04
	[Gender=.00]	1.88*	6.56
	[Gender=1.00]	0b	.
	[Ethnicity=.00]	0.61	1.84
	[Ethnicity=1.00]	0b	.
	[Marital_status=.00]	-0.12	0.88
	[Marital_status=1.00]	0b	.
	[Education=.00]	-1.35	0.26
	[Education=1.00]	-0.28	0.76
	[Education=2.00]	0b	.
	[Occupation=.00]	-0.59	0.56
	[Occupation=1.00]	-0.16	0.85
	[Occupation=2.00]	0b	.
	[Physical_activity=.00]	-1.10*	0.33
	[Physical_activity=1.00]	0b	.

a. The reference category is Non-Smoker.
b. Note: **P-value$<$0.05*
c. The notation “0b” indicates that the corresponding variable is used as the reference category for comparison.

4 CONCLUSION

The primary objectives of this study were to determine how different demographic factors are associated with smoking status (non-smoker, ex-smoker, daily smoker). The multinomial logistic regression analysis illustrates how demographic factors like weight, height, gender, ethnicity, marital status, education level, occupation, and physical activity are associated with the likelihood of being in one of these smoking categories compared to a reference category (non-smoker). The results indicated that some demographic factors are significant predictors of smoking status. There is a higher proportion of males who are daily smokers as compared to females.

Moreover, individuals from the Malay ethnic group show a higher probability of being a daily smoker. On the other hand, education was an important variable, where higher education significantly reduced the chances of smoking. This also underlines the importance of education in smoking prevention and cessation efforts. By identifying these key predictors, public health officials and policymakers can develop more effective strategies to specific populations to reduce smoking rates and improve public health outcomes.

However, the goodness-of-fit test shows that the model does not fit, which is indicative that the model is not good enough to reflect the relationship between the predictors and the outcome variable. This poor fit may affect its predictive performance. Further, it would be worthwhile to investigate advanced models in the future, which would analyze a broader set of demographic aspects, including but not limited to the income levels or the affordability of healthcare. Future research should also consider collecting primary data with approximately equal and sufficiently large samples for each category of smoking status for better inferences on these relationships.

REFERENCES

[1] M. B. Reitsma, L. S. Flor, E. C. Mullany, V. Gupta, S. I. Hay, and E. Gakidou, "Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019," *Lancet Public Health*, vol. 6, no. 7, 2021, doi: 10.1016/S2468-2667(21)00102-X.

[2] J. A. Fidler, R. West, C. H. M. Van Jaarsveld, M. J. Jarvis, and J. Wardle, "Does smoking in adolescence affect body mass index, waist or height? Findings from a longitudinal study," *Addiction*, vol. 102, no. 9, 2007, doi: 10.1111/j.1360-0443.2007.01910.x.

[3] Q. Wang, "Smoking and body weight: Evidence from China health and nutrition survey," *BMC Public Health*, vol. 15, no. 1, 2015, doi: 10.1186/s12889-015-2549-9.

[4] S. Kakoei, A. H. Nekouei, S. Kakooei, and H. Najafipour, "The effect of demographic characteristics on the relationship between smoking and dry mouth in Iran: A cross-sectional, case-control study," *Epidemiol Health*, vol. 43, 2021, doi: 10.4178/EPIH.E2021017.

[5] S. Khadanga, B. Yant, P. D. Savage, J. Rengo, and D. E. Gaalema, "Objective measure of smoking status highlights disparities by sex," *American Heart Journal Plus: Cardiology Research and Practice*, vol. 17, 2022, doi: 10.1016/j.ahjo.2022.100171.

[6] A. A. Zainel *et al.*, "Investigating the socio-demographic characteristics and smoking cessation incidence among smokers accessing smoking cessation services in primary care settings of Qatar, a Historical Cohort Study," *Discover Public Health*, vol. 21, no. 1, p. 3, 2024, doi: 10.1186/s12982-024-00124-x.

[7] J. S. Rose, L. Chassin, C. C. Presson, and S. J. Sherman, "Demographic factors in adult smoking status: Mediating and moderating influences," *Psychology of Addictive Behaviors*, vol. 10, no. 1, 1996, doi: 10.1037/0893-164X.10.1.28.

[8] B. E. Garrett, B. N. Martell, R. S. Caraballo, and B. A. King, "Socioeconomic differences in cigarette smoking among sociodemographic groups," *Prev Chronic Dis*, vol. 16, no. 6, 2019, doi: 10.5888/pcd16.180553.

[9] W. S. Chung, P. T. Kung, H. Y. Chang, and W. C. Tsai, "Demographics and medical disorders associated with smoking: a population-based study," *BMC Public Health*, vol. 20, no. 1, 2020, doi: 10.1186/s12889-020-08858-4.

[10] E. Raptou and G. Papastefanou, "An empirical investigation of the impact of smoking on body weight using an endogenous treatment effects model approach: The role of food consumption patterns," *Nutr J*, vol. 17, no. 1, 2018, doi: 10.1186/s12937-018-0408-0.

[11] D. C. Seo, N. Jiang, and L. J. Kolbe, "Association of smoking with body weight in us high school students, 1999-2005," *Am J Health Behav*, vol. 33, no. 2, 2009, doi: 10.5993/AJHB.33.2.9.

[12] J. Yoon and S. L. Bernell, "Link between perceived body weight and smoking behavior among adolescents," *Nicotine and Tobacco Research*, vol. 18, no. 11, 2016, doi: 10.1093/ntr/ntw116.

[13] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Ann Card Anaesth*, vol. 22, no. 1, 2019, doi: 10.4103/aca.ACA_157_18.

[14] H. Abdul Hamid, Y. Bee Wah, X. J. Xie, and O. Seng Huat, "Investigating the power of goodness-of-fit tests for multinomial logistic regression," *Commun Stat Simul Comput*, vol. 47, no. 4, 2018, doi: 10.1080/03610918.2017.1303727.

[15] D. Pregibon, "Logistic Regression Diagnostics," *The Annals of Statistics*, vol. 9, no. 4, pp. 705–724, Jul. 1981, doi: 10.1214/aos/1176345513.

[16] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Newark, UNITED STATES: John Wiley & Sons, Incorporated, 2013. [Online]. Available: http://ebookcentral.proquest.com/lib/unimap-ebooks/detail.action?docID=1138225

[17] M. W. Fagerland, D. W. Hosmer, and A. M. Bofin, "Multinomial goodness-of-fit tests for logistic regression models," *Stat Med*, vol. 27, no. 21, 2008, doi: 10.1002/sim.3202.

[18] R. W. Cooksey, "Descriptive Statistics for Summarising Data," in *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, 2020. doi: 10.1007/978-981-15-2537-7_5.

[19] J. F. Hair Jr., R. E. Anderson, B. J. Babin, and W. C. Black, "Multivariate Data Analysis, Multivariate Data Analysis," *Book*, vol. 87, no. 4, 2019.

[20] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," 1991. doi: 10.1093/biomet/78.3.691.