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

EKST

Vol. 6 No. 1 (2026) 512-522
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Analysis of Social Anxiety Factors using Ordinal Logistic Regression in Australia 2019-2022

Lee Kar Wen[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.2026.06.01.057

Article Info

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

Keywords

Social Anxiety Disorder, Cross Tabulation Analysis, Chi-Square Test, Ordinal Logistic Regression

Abstract

Social Anxiety Disorder (SAD) is a prevalent mental health condition that is characterised by intense fear of social interactions and negative evaluation. It's a global prevalence that has increased in recent years, where previous studies focused on limited factors or used binary classifications, resulting in an overlook of the gradual severity of anxiety. This study aims to determine the main characteristics of SAD, identify the associations between anxiety levels and explanatory variables and estimate the likelihood of individuals experiencing varying levels of social anxiety. This study utilised a Social Anxiety Dataset comprising 11,000 individuals in Australia between 2019 and 2022, employing Cross-Tabulation Analysis (CTA), the Chi Square Test, and Ordinal Logistic Regression (OLR) models to achieve the research objectives. The first objective used a CTA to reveal main characteristics with higher anxiety levels associated with shorter sleep duration, lower physical activity, abnormal breathing rate, more frequent therapy frequency, poor diet quality, high stress, sweating, dizziness, and family history of anxiety. These associations have been assessed using the Chi-Square Test, with the results showing that all explanatory variables are significantly associated with anxiety levels (p-value is less than 0.05), except for gender. Meanwhile, the OLR models are developed to estimate the likelihood of individuals experiencing different anxiety levels. After comparing several model specifications, Model 2, which excludes gender, performs best according to the statistical metric. Model 2 records the lowest Akaike Information Criterion (AIC) with a value of 33174.83, the lowest corrected Akaike Information Criterion (AICc) at 33175.10, the lowest Bayesian Information Criterion (BIC) at 33452.44, deviance of 33098.83 and McFadden's Pseudo of 0.2465 indicating that Model 2 demonstrates a strong overall fit. The model 2 revealed that risk factors increase the likelihood of SAD severity with an odds ratio greater than one, including excessive caffeine intake at 2.6585, excessive alcohol consumption at 1.1279, elevated stress levels at 1.8898, higher sweating levels at 1.0529, smoking at 1.1133, dizziness at 1.086, family history of anxiety at 1.439, medication use at 1.1549, and being a freelancer at 1.2327. The findings highlight the importance of lifestyle, physiological, and behavioural factors in shaping anxiety severity and demonstrate the value of ordinal logistic regression in mental health research.

1. Introduction

Social Anxiety Disorder (SAD) is a common and debilitating mental health condition characterised by severe and persistent fear of one or more social situations in which one is likely to be criticised by others as defined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [1]. SAD is associated with substantial impairments in social, academic, and occupational functioning and often leads to long-term avoidance behaviours that negatively affect daily social interactions, thereby hindering personal growth, professional development, and overall quality of life [2].

Global and regional evidence highlights the growing burden of SAD, particularly among adolescents and young adults, with Malaysian studies reporting anxiety-related symptoms in approximately 40% of school-going adolescents [3]. Despite its high prevalence, SAD remains underdiagnosed and inadequately managed in primary care, with fewer than half of affected individuals receiving appropriate treatment [4].

A key limitation in prior research on the factor's driving SAD is the frequent use of binary logistic regression or simple cross-sectional analysis, which divides social anxiety into simple presence or absence categories and ignores the ordered pattern of severity levels. This binary approach is a significant limitation, as SAD is clinically conceptualized and assessed as a dimensional construct with a spectrum of severity (mild, moderate, severe), information that is lost when the outcome is dichotomize. Otherwise, some of the research concentrates on specific factors of social anxiety, which means the analysis does not cover a comprehensive range of social anxiety disorder factors. This study employs a dual-method approach, combining OLR and Cross Tabulation Analysis (CTA), to overcome this limitation. Unlike binary models, OLR can determine whether an independent variable increases the odds of being in a higher severity category providing more clinically actionable insights. Furthermore, preliminary use of OLR in related anxiety research has successfully identified severity-specific risk gradients, such as the escalating impact of childhood trauma on the severity of generalized anxiety, demonstrating its established utility for modelling anxiety spectra [5].

The ordinal logistic regression (OLR) model is widely used in medical and epidemiologic research. The ordinal response variables represent naturally ordered outcomes such as pain severity or clinical event hierarchies It is often applied to rank-transformed continuous data to ensure robust inferences. A key advantage of ordinal logistic models is that they use only the rank order of outcomes without assuming equal spacing, making them well-suited for analysing ordered data in medical research [6]. In particular, the Proportional Odds (PO) and Continuation Ratio (CR) models are the standard modelling frameworks for OLR models.

This study aims to explore the main characteristics of SAD through CTA. Subsequently, the study will examine the potential association between different levels of anxiety and use the chi-square test to analyse the data. Finally, OLR will be applied to estimate the likelihood of individuals experiencing varying levels of social anxiety.

2. Method

2.1 Data Description

The dataset, known as the "Social Anxiety Dataset," was obtained from the Kaggle website, which originally sourced it from the Department of Health, Government of Western Australia. The data collection was conducted in Australia between 2019 and 2022 [7]. Table 1 shows the dataset that was compiled from real-world survey responses and observational studies on behavioural and psychological factors associated with social anxiety. The dataset comprises 11,000 samples of individuals and 19 variables, which vary in demographics, health backgrounds, and life experiences, contributing to the development of social anxiety scale from 1 to 10. Demographic and background profile variables include age, gender, occupation, family history of anxiety, and recent major life events. Variables related to lifestyle and habits include sleep hours, physical activity, diet quality, alcohol use, caffeine intake, and smoking habits. Psychological factors are comprehensively represented and include heart rate, breathing rate, stress level, sweating level, dizziness, medication use, and frequency of therapy. The study includes both categorical and numerical variables.

Table 1 The description of variables in dataset

Variable	Category	Description	Level of Measurement	Variable Type
1. Age	Demographic and background profile	18, …, 64	Discrete	Numerical

2.	Gender	Demographic and background profile	Female, Male, Other	Nominal	Categorical
3.	Occupation	Demographic and background profile	Artist, Nurse, Scientist, ..., Other	Nominal	Categorical
4.	Sleep hours	Lifestyle and Habit	2.3, ..., 11.3	Continuous	Numerical
5.	Physical activity	Lifestyle and Habit	0, ..., 10.1	Continuous	Numerical
6.	Diet quality	Lifestyle and Habit	1-10	Ordinal	Categorical
7.	Alcohol use	Lifestyle and Habit	0, ..., 19	Discrete	Numerical
8.	Caffeine intake	Lifestyle and Habit	0, ..., 599	Discrete	Numerical
9.	Smoking Habit	Lifestyle and Habit	Yes, No	Nominal	Categorical
10.	Heart rate	Psychological Factor	60, ..., 119	Continuous	Numerical
11.	Breathing rate	Psychological Factor	12, ..., 29	Continuous	Numerical
12.	Stress level	Psychological Factor	1-10	Ordinal	Categorical
13.	Sweating level	Psychological Factor	1-5	Ordinal	Categorical
14.	Dizziness	Psychological Factor	Yes, No	Nominal	Categorical
15.	Family history of anxiety	Demographic and background profile	Yes, No	Nominal	Categorical
16.	Use of medication	Psychological Factor	Yes, No	Nominal	Categorical
17.	Therapy frequency	Psychological Factor	0,1,2,3...	Discrete	Numerical
18.	Recent major life event	Demographic and background profile	Yes, No	Nominal	Categorical
19.	Anxiety Level	Response variable	1-10	Ordinal	Categorical

The categorical variables describe qualitative characteristics of the participants and are divided into nominal and ordinal types. Nominal categorical variables include gender, occupation, smoking habit, dizziness, family history of anxiety, use of medication, and recent major life events. Ordinal categorical variables include diet quality, stress level, sweating level, and anxiety level. Anxiety level is treated as the response variable in this study and is measured on a scale from 1 to 10, representing increasing levels of anxiety from low to severe. The numerical variables represent quantitative measurements and are classified as discrete or continuous. Discrete numerical variables include age, alcohol use, caffeine intake, and therapy frequency. Continuous numerical variables include sleep hours, physical activity, heart rate, and breathing rate. The dataset, which is diverse in terms of various measurement levels and different aspects of life, will assist future analysis.

2.2 Cross Tabulation Analysis

Cross Tabulation Analysis (CTA), also known as a contingency table, is used to investigate relationships between categorical variables by organising data into contingency tables. The rows represent categories of one variable, the columns represent categories of another variable, and each cell shows the frequency of a specific category combination [8]. This method enables the testing of statistical independence using the chi-square test, and graphical displays can further support the interpretation of results. In this study, CTA is used to determine the main characteristics of Social Anxiety Disorder (SAD).

2.3 Chi-Square Test

Chi-Square Test for Independence is applied in Cross Tabulation Analysis (CTA) to assess whether a significant relationship exists between two categorical variables. The hypotheses for the Chi-Square Test for Independence were formulated as a null hypothesis and an alternative hypothesis. The null hypothesis states that there is no association between the row and column variables, meaning they are statistically independent of each other. In contrast, the alternative hypothesis states that there is a significant association between the row and column variables, indicating that the variables are dependent on each other.
The Chi-Square test is calculated by Equation (1) based on [9];

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

where:

O is the observed frequency,
E is the expected value.

The significance level for the Chi-Square test is set at 0.05. The decision criteria will state to reject the null hypothesis if the Chi-Square statistic exceeds the critical value from the Chi-Square distribution with $(r-1)(c-1)$ degrees of freedom, where r is the number of rows and c is the number of columns.

2.4 Ordinal Logistic Regression Framework

Ordinal Logistic Regression (OLR) is a form of Multinomial Logistic Regression used when the dependent variable has more than two ordered categories. The Proportional Odds (PO) model is a standard OLR method suitable for ordered outcomes such as satisfaction or severity levels. The PO model can be expressed as in Equation (2) based on [6];

$$\Pr\left[Y \geq j \mid X\right] = \frac{1}{1+\exp\left[-\left(\alpha_j + X\beta\right)\right]} \tag{2}$$

where:

$j = 1, 2, \ldots, k$
α_j is a threshold-specific intercept or cut-point,
β is the common vector of regression coefficients for the predictor variables.

2.4.1 Model Assumptions

Ordinal Logistic Regression (OLR) assumes that the response variable is ordinal and that the effect of each predictor is consistent across all levels of the outcome. It also assumes no multicollinearity among predictors, which is commonly checked using the Variance Inflation Factor (VIF) [10]. For models that include categorical predictors alongside continuous ones, the Generalized Variance Inflation Factor (GVIF) is the appropriate measure, as it accounts for the multiple degrees of freedom associated with categorical variables and provides a comparable indicator of collinearity [11]. For continuous predictors, a linear relationship with the log odds is assumed; observations must be independent, and an adequate sample size is required. These assumptions can be evaluated using tests such as the Brant test and graphical diagnostics [12].

2.4.2 Model Evaluation

Model evaluation in ordinal logistic regression focuses on assessing the model's fit and explanatory power, as well as comparing competing models to determine the best fit. Standard measures include AIC, BIC, deviance, and McFadden's Pseudo-R^2. The Akaike Information Criterion (AIC), developed by Akaike in 1973, is defined as Equation (3) according to [13];

$$AIC = 2p - 2\ln(L) \tag{3}$$

where p is the number of parameters and L is the maximum likelihood of the model. Corrected Akaike Information Criterion (AICc) includes an additional small-sample bias correction term that makes it more accurate and reliable across a wider range of sample sizes. The AICc introduces an additional bias-correction term for finite sample sizes, yielding the formula [14];

$$AICc = -2\ln(L) + 2p + \frac{2p(p+1)}{n-p-1} \tag{4}$$

The Bayesian Information Criterion (BIC) which was presented by Schwarz in 1978 is calculated as in Equation (5) according to [15];

$$BIC = p\ln n - 2\ln L \tag{5}$$

where:
p is the number of parameters,
n is the sample size used for model estimation,
L is the maximum likelihood of the model.

AIC does not indicate how well a model fits the data in absolute terms but helps compare competing models. BIC, on the other hand, penalises additional parameters more severely than AIC. This frequently results in BIC being more cautious than AIC when selecting models. The lower values for AIC and BIC indicate a better model. Model explanatory power is assessed using McFadden's-R-squared, defined as Equation (6) according to [13];

$$R^2 = 1 - \frac{\ln L_{fitted}}{\ln L_{null}} \tag{6}$$

where:
$\ln L_{fitted}$ is the log-likelihood of the fitted model,
$\ln L_{null}$ is the log-likelihood of the null model.

McFadden's Pseudo-R^2 values typically range from 0 to 1 with values closer to 1 indicating better model performance. A value above 0.2 is generally interpreted as a strong model fit. Goodness of fit is further evaluated using the deviance statistic. A minor deviance relative to its degrees of freedom indicates an adequate model fit. The deviance can be calculated as Equation (7) according to [16];

$$\text{Deviance} = -2\left(\log L_{fitted} - \log L_{saturated}\right) \tag{7}$$

where:
$\log L_{fitted}$ is the log-likelihood of the fitted model,
$\log L_{saturated}$ is the log-likelihood of the saturated model.

3. Result and Discussion

3.1 Data Preprocessing

Prior to analysis, the dataset was screened to assess completeness, accuracy, and consistency, including checks for missing values, duplicate records, outliers, and implausible responses. Records with incomplete data were excluded, resulting in a final dataset with no missing values. To meet the requirement of cross tabulation analysis and the chi-square test, numerical variables were categorised into categorical variables based on authoritative research and established health guidelines. As presented in Table 2, sleep duration was categorised as sufficient (≥7 hours per night) or insufficient (<7 hours per night) [17]. Physical activity was classified as sufficient when participants engaged in at least 150 minutes (2.5 hours) of activity per week and insufficient when activity levels were below this recommendation [18]. Caffeine intake was categorised as acceptable when daily consumption did not exceed 400 milligrams and excessive when intake exceeded this threshold [19]. Alcohol consumption was classified as acceptable for participants consuming 10 standard drinks or fewer per week and excessive for those consuming more than 10 standard drinks per week [20]. Heart rate was categorised as normal (60–100 beats per minute) or abnormal (below 60 or above 100 beats per minute) [21], while breathing rate was classified as normal (12–20 breaths per minute) or abnormal (less than 12 or greater than 20 breaths per minute) [22]. Age was classified into three groups: young adults (18–39 years), middle-aged adults (40-59 years), and older adults (60 years old and above). [23] Therapy frequency was categorised into low (0–1 session per month), moderate (2–4 sessions per month), and high (more than 4 sessions per month) [24].

Table 2 The categorisation of numerical variables

Variables	Categorical	Description
Sleep Hours	Sufficient	Respondent who having 7 or more hours of sleep
	Insufficient	Respondent who having less than 7 hours of sleep
Physical Activity	Sufficient	Respondent who having at least 150 minutes (2.5hours) per week
	Insufficient	Respondent who having less than 150 minutes (2.5hours) per week
Caffeine Intake	Acceptable	Respondent who having not exceed 400 milligrams of caffeine
	Excessive	Respondent who having exceed 400 milligrams of caffeine
Alcohol Use	Acceptable	Respondent who drinking 10 standard drinks or fewer per week
	Excessive	Respondent who drinking more than 10 standard drinks per week
Heart Rate	Normal	Respondent who having 60 to 100 beats per minute

	Abnormal	Respondent who having below 60 beats per minute or above 100 beats per minute
Breathing Rate	Normal	Respondent who having 12 to 20 breaths per minute
	Abnormal	Respondent who having less than 12 breaths per minute or greater than 20 breaths per minute
Age	Young Adult	Respondent who between 18 and 39 years old
	Middle-Aged Adult	Respondent who between 40 and 59 years old
	Older Adult	Respondent who 60 years old and above
Therapy Frequency	Low	Respondent who having 0 and 1 session per month
	Moderate	Respondent who having 2 to 4 sessions per month
	High	Respondent who having more than 4 sessions per months

3.2 Cross Tabulation Analysis

After completing data preprocessing and transforming numerical variables into categorical forms, Cross Tabulation Analysis (CTA), together with chi-square testing, was applied to evaluate the relationships between social anxiety levels and the explanatory variables.

CTA showed that most respondents are in the mid-range of social anxiety levels. Across all variables, social anxiety levels were predominantly concentrated in the moderate range (levels 3–4), indicating that moderate symptom severity was most common within the sample. Among psychosocial and lifestyle factors, stress level exhibited the strongest and most consistent association with social anxiety severity. Participants reporting higher stress levels (levels 8–10) demonstrated markedly higher frequencies at severe social anxiety levels (8–10), indicating a clear positive relationship. Diet quality also displayed a notable association with SAD. Lower diet quality was associated with increased frequencies of severe social anxiety, whereas higher diet quality showed little to no representation at extreme anxiety levels (9–10). Sleep duration and physical activity showed similarly pronounced patterns, with individuals reporting insufficient sleep (<7 hours) and insufficient physical activity (<150 minutes per week) disproportionately represented at higher anxiety levels. These findings are consistent with previous research indicating that inadequate sleep and low physical activity are associated with increased risk of social anxiety [25].

In contrast, occupation and gender exhibited relatively uniform distributions across social anxiety levels, suggesting no substantial association with SAD severity. With respect to physiological and symptom-related variables, elevated sweating levels, the presence of dizziness, and abnormal breathing rates were consistently associated with higher social anxiety levels. Participants exhibiting these physiological indicators were more frequently represented in the severe anxiety categories. These findings align with prior studies identifying abnormal breathing, excessive sweating, and dizziness as common symptoms of social anxiety disorder [27][28]. Additionally, individuals with a family history of anxiety demonstrated a greater concentration at higher social anxiety levels, supporting a potential familial or genetic influence. Variables related to substance use, including smoking, caffeine intake, and alcohol consumption, showed moderate associations with social anxiety severity. While anxiety levels in both groups clustered around the moderate range, participants reporting smoking or excessive caffeine or alcohol intake were relatively more represented at severe anxiety levels. Finally, therapy frequency demonstrated a clear gradient with social anxiety severity. Participants attending therapy more frequently (>4 sessions per month) were predominantly observed at higher anxiety levels, whereas those with low or moderate therapy frequency were mainly concentrated within mild to moderate anxiety ranges.

Overall, the CTA findings indicate that respondents with severe social anxiety are characterised by poor diet quality, high stress levels, frequent or intense sweating, dizziness, family history of anxiety, insufficient sleep duration, insufficient physical activity, and abnormal breathing rates.

3.3 Chi-Square Test

The chi-square test has been conducted to determine the association between exploratory variables and anxiety levels. The results of the chi-squared test will be tabulated in Table 3 to summarise all the exploratory variables.

Table 3 Cross tabulation of anxiety level with exploratory variables with chi-square test

Variable	***p*-value**	**Variable**	***p*-value**
Gender	> 0.1	Age	<0.05
Occupation	< 0.05	Sleep hours	<0.05
Diet quality	< 0.05	Physical activity	<0.05
Smoking habit	< 0.05	Alcohol use	<0.05

Stress level	< 0.05	Caffeine intake	<0.05
Sweating level	< 0.05	Heart rate	<0.05
Dizziness	< 0.05	Breathing rate	<0.05
Family history of anxiety	< 0.05	Therapy frequency	<0.05
Use of medication	< 0.05	Recent major life event	< 0.05

The analysis indicates that most variables demonstrate statistically significant associations with anxiety levels, with p-values smaller than 0.05 for all variables except gender. For gender, the p-value is greater than 0.1, indicating no statistically significant association with social anxiety levels. This finding is consistent with prior research suggesting that gender differences in SAD are not always substantial [26][27].

3.4 Ordinal Logistic Regression

This section develops and evaluates ordinal logistic regression models using the Proportional Odds framework to identify key determinants of social anxiety levels, measured on a 1-10 scale. An initial complete model, Model 1, including all predictors, was fitted, followed by assumption checking using Generalized Variance Inflation Factor (GVIF) and the Brant Wald test. Results indicate no multicollinearity, with the GVIF of all variables remaining less than 5. The overall proportional odds assumption holds, with a p-value greater than 0.05; however, some individual predictors exhibit violations with p-values less than 0.05.

Based on these findings and the Chi-square test results indicating that gender was not significantly associated with social anxiety levels, a reduced model, Model 2 was subsequently constructed, excluding gender as a predictor. These models are developed to evaluate the robustness of associations and to identify the best model that adequately explains variation in social anxiety levels. The equations for Models 1 and 2 show as Equations (7) and (8) respectively, which vary only in the predictors included. In the model equations, j represents the cumulative threshold for the ordinal outcome, corresponding to social anxiety levels 1 through 9; each θ_j denotes the log-odds of a respondent's social anxiety being at or below level j, while the effect of each predictor (β_k) is assumed constant across all thresholds under the proportional odds assumption.

$$\begin{aligned} Logit\left[P\left(Y \leq j\right)\right] = \theta_j - \beta_1 Age - \beta_2 Sleep\ hours - \beta_3 Physics\ activity - \beta_4 Caffeine\ intake - \beta_5 Alcohol\ use - \\ \beta_6 Heart\ rate - \beta_7 Breathing\ rate - \beta_8 Therapy\ frequency - \beta_9 Occupation - \beta_{10} Diet\ quality - \\ \beta_{11} Stress\ level - \beta_{12} Sweating\ level - \beta_{13} Smoking - \beta_{14} Dizziness - \beta_{15}\ Family\ history\ of\ anxiety - \\ \beta_{16} Use\ of\ medication - \beta_{17} Gender - \beta_{18} Recent\ major\ life\ event \end{aligned} \tag{7}$$

$$\begin{aligned} Logit\left[P\left(Y \leq j\right)\right] = \theta_j - \beta_1 Age - \beta_2 Sleep\ hours - \beta_3 Physics\ activity - \beta_4 Caffeine\ intake - \beta_5 Alcohol\ use - \\ \beta_6 Heart\ rate - \beta_7 Breathing\ rate - \beta_8 Therapy\ frequency - \beta_9 Occupation - \beta_{10} Diet\ quality - \\ \beta_{11} Stress\ level - \beta_{12} Sweating\ level - \beta_{13} Smoking - \beta_{14} Dizziness - \beta_{15}\ Family\ history\ of\ anxiety - \\ \beta_{16} Use\ of\ medication - \beta_{18} Recent\ major\ life\ event \end{aligned} \tag{8}$$

Table 4 presents a comparison of two models based on model fit statistics, including AIC, AICc, BIC, deviance, and McFadden's pseudo-R^2. Both Model 1, which includes all variables, and Model 2, which excludes gender, show similar McFadden's pseudo-R-squared values of 0.2465, indicating that both models explain a similar proportion of the variation in social anxiety levels. Model 2 has slightly lower AIC and AICc values at 33174.83 and 33175.10 respectively, compared to 33177.10 for Model 1. Its Bayesian Information Criterion is also lower at 33452.44 compared to 33469.33 for Model 1, suggesting that Model 2 achieves a slightly better balance between model fit and simplicity. The deviance of Model 2 is 33098.83, slightly higher than Model 1 at 33097.10, but the difference is very small and not practically meaningful. Therefore, model 2 is selected as the best-fitting model for estimating the likelihood of social anxiety levels.

Table 4 Model Comparison of OLR with different variables

	AIC	AICc	BIC	Deviance	McFadden's Pseudo R^2
Model 1	33177.10	33177.10	33469.33	**33097.10**	0.2465
Model 2	**33174.83**	**33175.10**	**33452.44**	33098.83	0.2465

Table 5 presents the odds ratios, confidence intervals, and *p*-values for the Model 2. An ordinal logistic regression was conducted to examine factors associated with social anxiety levels (1–10). In this model, the thresholds (cut-points) represent the boundaries between adjacent categories of social anxiety, with the notation 1|2, 2|3, ..., 9|10 referring to the separation between category 1 and 2, category 2 and 3, and so on up to category 9 and 10. Each threshold corresponds to the odds of being at or below a given category versus being in a higher category. The estimated thresholds increased monotonically from 0.2212 (1|2) to 4439.0071 (9|10), all statistically significant with p-values less than 0.05, indicating progressively higher odds of elevated social anxiety at higher levels. In the ordinal logistic regression model, an odds ratio (OR) less than 1 signifies a protective factor, indicating that the predictor is associated with reduced odds of being in a higher social anxiety level. Conversely, an odds ratio greater than 1 signifies a risk factor, indicating that the predictor is associated with increased odds of being in a higher social anxiety level, when all other variables in the model are held constant.

Table 5 Odds Ratio for Model 2

Variables	Odd Ratio (OR)	CI lower	CI upper	*p*-value	Interpretation
1|2	0.2212	0.1727	0.2833	< 0.001	Cut-point between level 1 and ≥2
2|3	1.2671	0.9901	1.6215	< 0.001	Cut-point between level ≤2 and ≥3
3|4	7.0916	5.5080	9.1304	< 0.001	Cut-point between level ≤3 and ≥4
4|5	38.9210	29.9678	50.5490	< 0.001	Cut-point between level ≤4 and ≥5
5|6	182.5045	139.4922	238.7798	< 0.001	Cut-point between level ≤5 and ≥6
6|7	485.8074	369.8512	638.1183	< 0.001	Cut-point between level ≤6 and ≥7
7|8	643.7251	489.5584	846.4405	< 0.001	Cut-point between level ≤7 and ≥8
8|9	1601.1111	1211.8141	2115.4703	< 0.001	Cut-point between level ≤8 and ≥9
9|10	4439.0071	3320.8885	5933.5879	< 0.001	Cut-point between level ≤9 and 10
Age Group; Middle-Aged Adults	1.0801	0.9884	1.1804	< 0.1	Not significant
Age Group: Older Adults	0.9162	0.8383	1.0014	< 0.1	Not significant
Sleep Hours; Sufficient	0.3797	0.3525	0.4089	< 0.001	Reduces odds by 62%
Physical Activity; Sufficient	0.6755	0.6290	0.7255	< 0.001	Reduces odds by 32%
Caffeine Intake; Excessive	2.6585	2.3847	2.9638	< 0.001	Increases odds by 166%
Alcohol Use; Excessive	1.1279	1.0531	1.2080	< 0.001	Increases odds by 13%
Heart Rate; Normal	0.8775	0.8160	0.9436	< 0.001	Reduces odds by 12%
Breathing Rate; Normal	0.8553	0.7976	0.9171	< 0.001	Reduces odds by 14%
Therapy.Sessions; Normal	0.2299	0.2056	0.2571	< 0.001	Reduces odds by 77%
OccupationAthlete	0.9994	0.8404	1.1885	> 0.1	Not significant
OccupationChef	1.0839	0.9128	1.2871	> 0.1	Not significant

OccupationDoctor	0.9933	0.8284	1.1910	> 0.1	Not significant
OccupationEngineer	0.9972	0.8314	1.1960	> 0.1	Not significant
OccupationFreelancer	1.2327	1.0370	1.4653	< 0.01	Increases odds by 23%
OccupationLawyer	1.0843	0.9036	1.3012	> 0.1	Not significant
OccupationMusician	0.9184	0.7752	1.0880	> 0.1	Not significant
OccupationNurse	1.0141	0.8538	1.2045	> 0.1	Not significant
OccupationOther	0.9132	0.7679	1.0860	> 0.1	Not significant
OccupationScientist	1.1313	0.9433	1.3569	> 0.1	Not significant
OccupationStudent	1.0359	0.8736	1.2283	> 0.1	Not significant
OccupationTeacher	0.9121	0.7667	1.2283	> 0.1	Not significant
Diet Quality 1-10	0.9561	0.9446	0.9678	< 0.001	Reduces odds by 4% per unit
Stress Level 1-10	1.8898	1.8591	1.9210	< 0.001	Increases odds by 89% per unit
Sweating Level 1-5	1.0529	1.0271	1.0794	< 0.001	Increases odds by 5% per unit
SmokingYes	1.1133	1.039	1.192	< 0.05	Increases odds by 11%
DizzinessYes	1.0855	1.0135	1.1627	< 0.05	Increases odds by 9%
Family History of Anxiety; Yes	1.4389	1.3382	1.5472	< 0.001	Increases odds by 44%
MedicationYes	1.1549	1.0786	1.2366	< 0.001	Increases odds by 15%
Recent Major Life Event; Yes	1.0492	0.9797	1.1236	> 0.1	Not significant

Several predictors were identified as protective factors against higher social anxiety levels (OR < 1, $p < 0.05$). Individuals reporting sufficient sleep (0.3797) had 62% lower odds of experiencing higher social anxiety, while adequate physical activity (0.676) reduced the odds by 32%. Similarly, normal heart rate (0.8775) and normal breathing rate (0.8553) were associated with reduced anxiety severity. Regular therapy sessions emerged as a strong protective factor (0.2299), lowering the odds of higher social anxiety by 77%. Additionally, higher diet quality demonstrated a cumulative protective effect, with each unit increase reducing the odds of higher anxiety by approximately 4% (0.9561).

In contrast, several variables functioned as risk factors (OR > 1, $p < 0.05$), significantly increasing the likelihood of higher social anxiety levels. Excessive caffeine intake substantially elevated risk (2.6585), more than doubling the odds of severe anxiety. Excessive alcohol consumption (1.1279) and elevated stress levels (OR = 1.8898 per unit increase) further intensified anxiety severity. Physiological and behavioural factors, including increased sweating levels (OR = 1.0529 per unit increase), smoking (1.1133), and dizziness (1.0855), were also associated with heightened anxiety. Moreover, individuals with a family history of anxiety faced significantly higher odds of severe anxiety (1.4389), highlighting genetic or environmental susceptibility. Medication use (1.1549) and being self-employed as a freelancer (1.2331) were likewise identified as risk factors.

Social anxiety severity is influenced by a combination of foundational lifestyle habits like sufficient sleep and regular exercise, targeted behaviours including caffeine intake, alcohol consumption, and stress management, access to professional care such as therapy, and inherent risk factors like family history. This condition also manifests in measurable physiological symptoms, for example changes in heart rate, breathing rate and sweating. For intervention, the most powerful and actionable strategies are engaging in therapy, ensuring sufficient sleep and physical activity, eliminating excessive caffeine intake and alcohol consumption, and proactively managing stress.

4. Conclusion and Recommendations

In conclusion, this study was successfully achieved the three research objectives. The cross-tabulation analysis identified main characteristics associated with social anxiety disorder. The key factors associated with higher anxiety include shorter sleep duration, lower physical activity, abnormal breathing rate, more frequent therapy, poor diet quality, high stress, sweating, dizziness, and family history of anxiety. The chi-square test confirmed that the explanatory variables were significantly associated with social anxiety levels, except for gender.

The ordinal logistic regression further refined these findings, with Model 2 which excludes gender, providing the best fit. Model 2 records the lowest Akaike Information Criterion (AIC) with a value of 33174.83, the lowest corrected Akaike Information Criterion (AICc) at 33175.10, the lowest Bayesian Information Criterion (BIC) at 33452.44, deviance of 33098.83 and McFadden's Pseudo of 0.2465. Model2 discovered that adequate sleep, sufficient physical activity, stable physiological indicators, and higher diet quality were associated with lower social anxiety, highlighting the role of healthy routines and physiological balance. Model 2 also revealed that risk factors with odds ratios greater than 1 included excessive caffeine intake (2.6585), excessive alcohol consumption (1.1279), elevated stress levels (1.8898), higher sweating levels (1.0529), smoking (1.1133), dizziness (1.0855), family history of anxiety (1.4389), medication use (1.1549), and being a freelancer (1.2327). Excessive caffeine intake and elevated stress were the strongest risk factors, while alcohol use, smoking, physiological symptoms, medication use, and a family history of anxiety collectively increased the likelihood of higher social anxiety severity. Overall, the results demonstrate that many factors influence social anxiety disorder, and the model provides a reliable way to estimate different levels of anxiety.

Despite these findings, one key limitation of this study is its reliance on self-reported secondary data, which may be subject to reporting bias and lacks detailed clinical assessments such as structured diagnostic interviews. As a result, the findings reflect observed associations rather than definitive causal relationships. To address this limitation, future research is recommended to incorporate clinically validated diagnostic tools or objective physiological measures alongside self-reported data. This approach would improve diagnostic accuracy and strengthen the interpretation of factors associated with social anxiety severity.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology at 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:*** Lee Kar Wen; ***solving the governing equation:*** Lee Kar Wen, Norziha Che Him; ***data collection:*** Lee Kar Wen; ***analysis and interpretation of results:*** Lee Kar Wen, Norziha Che Him; ***draft manuscript preparation:*** Lee Kar Wen, Norziha Che Him, Noor Azliza Abd Latif, Yusliandy Yusof. All authors reviewed the results and approved the final version of the manuscript.

References

[1] American Psychiatric Association. (2013). *Diagnostic And Statistical Manual Of Mental Disorders: DSM-5-TR* (5th ed., Vol. 21). 10.1097/NMD.0b013e3182a2168a

[2] Schneier, F., & Goldmark, J. (2015). Social anxiety disorder. In *Anxiety Disorders and Gender* (pp. 49–67). Springer International Publishing. https://doi.org/10.1007/978-3-319-13060-6_3

[3] Taufik, M. M. H., Rezali Muhammad Solihin, Shahein Nur Adilah, Sahril Nur Adilah, Sahril Norhafizah, Chan Ying Ying, & Wahab Nor'Ain Ab. (2022). 2022 Anxiety and Its Associated Factors among School-going Adolescents in Malaysia. *International Journal of Publec Health Research, 12*(2), 1608–1616.

[4] Chapdelaine, A., Carrier, J. D., Fournier, L., Duhoux, A., & Roberge, P. (2018). Treatment adequacy for social anxiety disorder in primary care patients. *PLoS One, 13*(11). https://doi.org/10.1371/journal.pone.0206357

[5] Bălțată, A., et al. (2021). "Statistical approaches to anxiety disorders: a systematic review." *Journal of Evidence-Based Psychotherapies, 21*(1), 119-142.

[6] Harrell, F. E. (2001). *Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis Second Edition Springer Series in Statistics* (2nd ed., Vol. 608). Springer. http://www.springer.com/series/692

[7] Monteiro, T. (2025). Social Anxiety Dataset. Kaggle.com. https://www.kaggle.com/datasets/natezhang123/social-anxiety-dataset/data

[8] Agresti, A., Franklin, C. A. ., & Klingenberg, Bernhard. (2018). Statistics : the art and science of learning from data (D. Lynch, Ed.; 4th ed.). Pearson Education Limited.

[9] Lyman Ott, R., & Longnecker, M. (2016). An Introduction to Statistical Methods & Data Analysis (7th ed.). Cengage Learning.Sukumar, G. (2024). AIC, BIC, Penalties and the Science behind Model selection. *Medium*.https://medium.com/%40gopal.sukumar/aic-bic-penalties-and-the-science-behind-model-selection-bd0d4d06d335

[10] Vatcheva, K. P., & Lee, M. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology*: Open Access, 06*(02). https://doi.org/10.4172/2161 1165.1000227
[11] Fox, J., & Weisberg, S. (2019). An R companion to applied regression (3rd ed.). Sage publications.
[12] Liu, A., He, H., Tu, X. M. ., & Tang, W. (2023). On testing proportional odds assumptions for proportional odds models. In General Psychiatry 36(3). BMJ Publishing Group. https://doi.org/10.1136/gpsych-2023-101048
[13] Gemechu, D. B., & Elifas, L. O. M. (2025). Ordinal Logistic Regression Model in Determining Factors Associated with Household Food Insecurity in Namibia. Open Journal of Statistics, 15(03), 272–293. https://doi.org/10.4236/ojs.2025.153014
[14] Nguyen, M. (2025) *8.4 Information Criteria for Model Selection | A Guide on Data Analysis*. Bookdown.org. https://bookdown.org/mike/data_analysis/information-criteria-for-model-selection.html
[15] Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. Briefings in Bioinformatics, 21(2), 553–565. https://doi.org/10.1093/bib/bbz016
[16] Agiakloglou, C., & Tsimpanos, A. (2023). Evaluating the performance of AIC and BIC for selecting spatial econometric models. *Journal of Spatial Econometrics*, *4*(1). https://doi.org/10.1007/s43071-022-00030-
[17] Yuan, W., Chen, L., Wu, Y., Su, B., Liu, J., Zhang, Y., Chen, M., Ma, Y., Guo, T., Wang, X., Ma, T., Ma, Q., Cui, M., Ma, J., & Dong, Y. (2023). Sleep time and quality associated with depression and social anxiety among children and adolescents aged 6–18 years, stratified by body composition. *Journal of Affective Disorders, 338*, 321–328. https://doi.org/10.1016/j.jad.2023.06.029 Affective Disorders, 338, 321–328.
[18] Physical activity. (n.d.). https://www.who.int/initiatives/behealthy/physical-activity
[19] Office of the Commissioner. (2024). Spilling the Beans: How Much Caffeine is Too Much? U.S. Food And Drug Administration. https://www.fda.gov/consumers/consumer-updates/spilling-beans-how-much-caffeine-too-much
[20] Health. (2019). How much alcohol is safe to drink? Australian Government Department of Health and Aged Care. https://www.health.gov.au/topics/alcohol/about-alcohol/how-much-alcohol-is-safe-to-drink?language=en
[21] Chourpiliadis, C., & Bhardwaj, A. (2022). Physiology, Respiratory Rate. National Library of Medicine; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK537306/
[22] Alvares, G. A., Quintana, D. S., Kemp, A. H., Van Zwieten, A., Balleine, B. W., Hickie, I. B., & Guastella, A. J. (2013). Reduced Heart Rate Variability in Social Anxiety Disorder: Associations with Gender and Symptom Severity. *PLoS One, 8*(7). https://doi.org/10.1371/journal.pone.0070468
[23] Horng, W.-B., Lee, C.-P., & Chen, C.-W. (2001). Classification of age groups based on facial features. Tamkang Journal of Science and Engineering, 4(3), 183–192.
[24] Fisher, J. (2024). Social Anxiety disorder: Treatments and Tips for Managing This Challenging Condition. Harvard Health; Harvard Health Publishing. https://www.health.harvard.edu/mind-and-mood/social-anxiety-disorder-treatments-and-tips-for-managing-this-challenging-condition
[25] Roh, K. J., Kim, M. K., Kim, J. H., & Son, E. J. (2018). Role of emotional distress in prolongation of dizziness: A cross-sectional study. *Journal of Audiology and Otology*, *22*(1), 6–12. https://doi.org/10.7874/jao.2017.00290
[26] McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, *45*(8), 1027–1035. https://doi.org/10.1016/j.jpsychires.2011.03.006
[27] Razidi, Umarudin, Yusuf, & Anuar. (2021). 2020 Investigating Gender Differences In Social Anxiety Disorder Among University Students. *The Asian Journal of Professional and Business Studies*, *2*(1). https://www.researchgate.net/publication/356776348