ISSN: 2317-0840

Academic stress determinant factors: an investigation according to Multinomial Logistic Regression

João Debastiani Neto[1†], Clicia Geovane Alves Pereira[1], Maria Graziela da Silva Fernandes[2], Ricardo Puziol de Oliveira[3]

[1] *Professor from the Science Department, State University of Maringá*
[2] *Mater's degree student of Applied and Computational Mathematics, State University of Londrina (UEL)*
[3] *Professor from the Statistics, State University of Maringá (UEM)*

Abstract: *The manifested stress in students represents a prevalent variable that focus in a substantial portion of the student demography. Many elements of psychological, physiological and academic nature can focus on the tension level experienced by the students. In order to conduct a statistical analysis on this issue, the current study is proposed, aiming to implement a Multinomial Logistic Model to identify factors and their impact on stress levels among nepalese students. The dataset was obtained from the Kaggle repository, named "StressLevelDataset", comprising 20 attributes assessing stress levels in 1100 students. These attributes were categorized into five factors: Psychological (PF), Physiological (PhF), Social (SF), Environmental (EF), and Academic (AF). For this investigation, stress analysis focused on PF, PhF, and FA. Key attributes within these factors include anxiety, self-esteem, depression, and history of mental illness (PF); headache, sleep quality, and respiratory problems (PhF); academic performance, study workload, student-teacher relationship, and concern about future career (AF). The adjusted multinomial model, using stress level as the response variable and the eleven covariates mentioned, yielded an accuracy of 89.09%. Considering these metrics, it was observed that the variables level of academic performance, level of sleep quality and level of study hours exerted the greatest influence in explaining the level of stress, in this order, with a relevance of 22.01%, 17.4% and 14.9%, respectively.*

Keywords: *Multinomial Distribution; Students; Regression Model; Levels of Education; Level of Stress.*

[†] Autor correspondente: jdneto@uem.br

Manuscrito recebido em: 07/06/2024
Manuscrito revisado em: 20/09/2024
Manuscrito aceito em: 23/09/2024

Introduction

The concept of stress between teenagers has received a considerable attention and highlights in the scientific community. According to the comments of Just and Enumo (2015), the adolescence represents a crucial phase in the life of young people, which is by characterized lots of transformations that cover not only physical and aspects, but also significant psychological dimensions. It is a period marked by transition which teenagers if find themselves constantly facing choices and decisions that aim the conquest of their necessary autonomy to build their futures (ELTINK; NUNES, 2020).

In that context, it is common that adolescence become an environment related to diverse pressures, resulting in meaningful emotional and psychological challenges lots of times. Anxiety, depression and stress are just some of the psychopathological manifestations that can emerge as a consequence of this scenario (MARTINS, *et al.*, 2021).

When a pedagogical perspective is considered, investigations like the conducted by Joyotsna and Amudha (2018), Pariat *et al.* (2014), Pitt *et al.* (2018) and Santos *et al.* (2017) highlight the growing concern with the level of pathologies developed in young people and teenagers. This restlessness has a foundation related to the fact that the environment educational, which ideally should be spaces for building knowledge, promoting lots of questins and the development of the scientific spirit, frequently reveal themselves as demotivators and highly stressful. This is, highly, because of the persistence in the traditional teaching approaches, which are reasoned in success and failure concepts related to assessments.

Several elements can contribute to the amplification of the stress level in students in many different educational stages. According to Pariat *et al.* (2014), these stressors can manifest themselves both internally and externally, being the economic, social and professional challenges the one which more exercise meaningful psychological pressure. In addition, an unsatisfactory academic developmente in works and exams can initiate a series of prejuditial elements to health, such as anxiety, substances abuse, depression and other psychological disorders.

At the what if refers to stresses, it is known this is pathology is classified in three different phases, outlined based in specific characteristics observed in the individuals: the alert phase, the resistance phase and the exhaustion phase (ABREU, *et al.*, 2002). The alert phase is characterized by the stressor perception and by the mobilization to face the situation, generating motivation and enthusiasm in the individual. In contrast, the resistance phase manifest itself when the stressor persists after the last phase, demanding a bigger energy mobilization to face it.

In that stage, symptoms can emerge such as tiredness, memory lapses, herpes and arterial pressure. Finally, the exhaustion phase is marked by the most severe effects to the individual health. Here, the symptoms start with a generalized tiredness, which evolves to insomnia, dermatological disorders, gastrointestinal and cardiovascular problems, emotional instability, and others (BORINI, *et al.*, 2015).

It is notable that there are several factors that can initiate stress in individuals, specially in students from diverse educational levels. Aiming to conduct a statistical analysis about it, this study propouses the Multinomial Logistic Model in three levels to identify the factors and its influence in the stress levels in Nepalese students. This approach allows a detailed analysis of the complex interactions that contribute to population stress. It is expected this study shows important results to the effective interventions and political development aiming to the well-being and the academic sucess.

Materials and Methods

Data set

This study used the programming language R to data analysis and interpretation, which the corpus derives from Kaggle repository, entitled ”StressLevelDataset”, using 20 attributes (covarietes) that instill in the 1100 Nepale students stress level.

These attributes were categorized in five fundamental factors: Psychological, Physiological, Social, Environmental and Academic. Considering the research nature, it was chosen to analyse the students stress according the Psychological, Physiological and Academic factors. The attributes that compound the mentioned factors are: anxiety (AN), self esteem (SF), depression (DP) and history of mental diseases (HD), which are related to the Psychological Factors; cephalea (CL), sleeping quality (SQ) and respiratory problems (RP), refer to the Physiological Factor; and finally, academic performance (AP), studyng workload (SW), relationship student-professor (SP) and concerning with the career future (CF), related to the Academic Factor. The Table 1 presents a variates summary as well as their natures.

Table 1: Varieties Description.

Variety	Description	Nature
HD	History of mental diseases	Binary
AN	Anxiety level	Categorical
SF	Self steem level	Categorical
SW	Studying workload level	Categorical
CL	Cephalea level	Categorical
DP	Depression level	Categorical
AP	Academic performance level	Categorical
ST	Stress Level	Categorical
CF	Concerning with the career future	Categorical
RP	Respiratory problems level	Categorical
SQ	Sleeping quality level	Categorical
SP	Relationship student-professor level	Categorical

Source: from the authors (2024).

Regarding the exploratory and descriptive statistical analyses, various approaches were employed to characterize the data and assess the performance of the adjusted model. Initially, bar charts were used to visualize the distribution of categorical variables, providing a clear understanding of the relative frequencies across different groups. To complement this visualization, frequency tables were constructed to detail the count and proportion of observations in each category, offering a more precise quantitative view.

Additionally, descriptive tables containing numerical information about the adjusted models were created. These tables included key metrics for performance evaluation, such as accuracy, sensitivity, and specificity. These metrics were crucial for assessing the quality of the models and their ability to correctly predict outcomes.

Considering that, this research aimed to implement a multinomial logistic model. It had as variety the Nepale students stress level, and, as varieties, the eleven attributes presented in Table 1.

Multinomial Logistic Regression

The multinomial logistic regression is a binary logistic regression generalization, which is modeled according to a categorical dependent variety Y with k. These are categories mutualy exclusive and a set of t independent varieties $x_1, x_2, ..., x_t$. In this regression model class, one of the k varieties categories is used as reference, aiming to compare with others categories. It is highlighted, yet, this reference category choice can happen in an aleatory way or by the researcher choice (FÁVERO, *et al.*, 2009).

So, the multinomial logistic regression model, with K categories, being chosen the 0 category as a reference to a set of t independent varieties $x_1, x_2, ..., x_t$, it can be expressed by:

$$g_k = ln\left[\frac{P(Y=k|x)}{P(Y=0|x)}\right] = \beta_{k0} + \sum_{j=1}^{t} \beta_{kj} x_j.$$

To study using a screen, consider that the variety depends on the existence of 3 categories (0 - low stress level, 1 - intermediary stress level, and 2 - high stress level) and being $x = (x_1, x_2, x_3, x_4, ..., x_{11})$ the vector of eleven covarieties from the model. The conditional probability function to the multinomial logistic regression to the k category rekated to the 0 reference category ? alert phase happens in the next logistic equation:

$$g_1(x) = \ln\left[\frac{P(Y=1|x)}{P(Y=0|x)}\right] = \beta_{10} + \sum_{j=1}^{11} \beta_{1j} x_j;$$

$$g_2(x) = \ln\left[\frac{P(Y=2|x)}{P(Y=0|x)}\right] = \beta_{20} + \sum_{j=1}^{11} \beta_{2j} x_j; \tag{1}$$

which k is the students stress level, $P(Y=k|x)$ is stress level conditional probability, being k the vector $\mathbf{x}$ that has the covarieties values; $P(Y=0|k)$ is the reference stress level conditional possibility gave by the vector $\mathbf{x}$. The equation 1 represents a multinomial regression expressed by the logit function (HOSMER, *et al.*, 2013).

According to Sofro *et al.* (2023), one of the technics used to esteem the multinomial logistic regression parameters is using maximum likelihood method, which the likelihood function is:

$$L(\beta) = \prod_{i=1}^{t} L_i(\beta) = \prod_{i=1}^{t} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \tag{2}$$

So, the log-likelihood function is:

$$L(\beta) = \sum_{i=1}^{t} \{y_i \cdot \ln \pi(x_i) + (1-y_i)\ln(1-\pi(x_i))\} \tag{3}$$

Aiming to steem the β parameter, it is seeked to maximize the log-likelihood function 3. This procedure can happen by the differentiation function 3 related to β and, after that, equaling it to zero, or using the iteration procedure from Newton-Raphson (SOFRO, *et al.*, 2023).

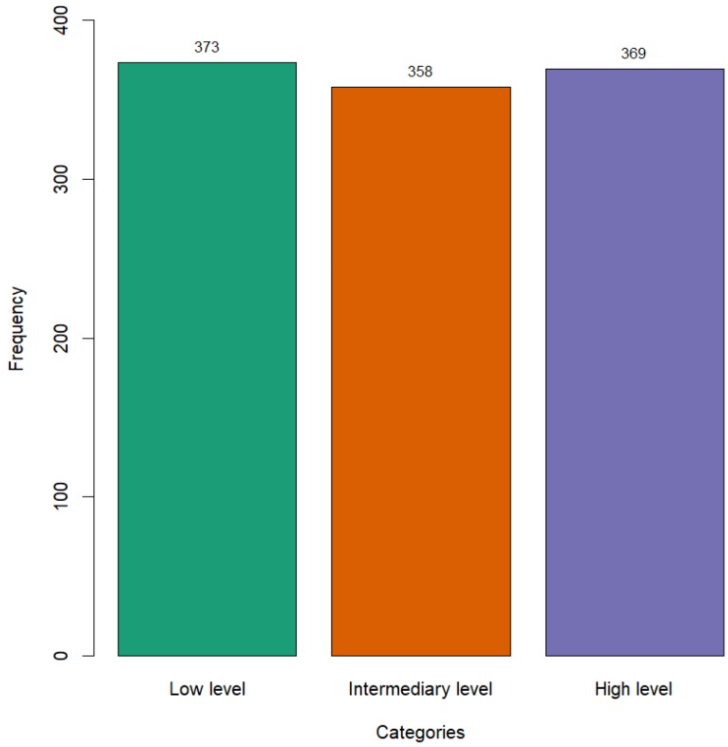
Results and discussion

Descriptive analysis and exploring the data

Aiming to stablish a multinomial logistic model in three leves to investigate the underlying factors and its influence in the different stress levels between the Nepale students, the data from a sample of 1100 participants was analyzed. This analysis was conducted considering eleven specific varieties, according to what is detailed in Table 1.

Initially, the model variable answer was analyzed, in other words, the students stress level. Therefore, the Figure 1 presents a bar graph of this variety according to the three categories

Figure 1: Bar graph - stress levels.



Source: from the authors (2024).

Table 2: Stress levels.

Stress	Frequency	Percentage (%)
Low	373	33.9
Intermediary	358	32.5
High	369	33.5

Source: from the authors (2024).

considered. Futhermore, the Table 2 shows the absolute frequency of these categories as well as their percentage according to the total number of students.

The results analysis presented in Figure 1 and in Table 2 revel there is no substantial difference between the frequencies in the three stress levels. It indicates a frequency equitable distribution of students in the three distincs stress levels considered. Considering the 1100 examined participants, it was observed that 373 individuals are classified in the low stress level, 369 in the high level, and 358 in the intermediary level. These values match, respectively, to 33.9%, 33.5% and 32.5% from the sample total. These data highlight a stress prevalence worrisome between the students, demonstrating that about one third of them face a stress level that meaningfully interfere in their daily lives, in physical and mental health (BUBLITZ, *et al.*, 2016).

Considering the inherent premises related to the multinomial logistic regression model, in the context of this investigation, the dependent variety was identified as the stress level in the students, which one of the categories was sent as a reference to make comparision with the others. Seeing the arbitrary character of this selection, it was chosen to designate a low level as a reference, like it was previously stipulated by De Lima *et al.* (2016).

Besides the previous considerations, there were conducted analysis related to the assumptions of the varieties in the data set, considering the multinomial logistic regression model application (Das Mangas, 2019). The dependent variety is related to the stress level of the students, presenting categories mutually exclusive, denoting that the students in a determined stress level cannot simultaneously belong to other categories. In addition, it was observed the independence in the observations, highlighting the repetition absence in the data set, which in each observation is from a different student.

Finally, information regarding multicollinearity between the covariates was investigated, that is, the possibility of a high correlation between the eleven independent variables of the model.

The Table 3 presents the variance inflation factor (VIF) values of the covariates considered. Through Table 3, there is evidence of the non-existence of multicollinearity between the independent variables, as indicated by the values of the covariate variance inflation factor, which did not exceed the limit of 3.14. This result is consistent with the literature, as described by Johnson and Wichern (1988), which establishes that only VIF values above 10 can raise concerns about the effects on the regression coefficients.

Table 3: Values of the variance inflation factors of the 11 covariates.

Covariety	Variance Inflation Factor
Anxiety level	3.023008
Self steem level	2.753082
Depression level	2.993747
History of mental diseases	2.132425
Cephalea level	2.404178
Sleeping quality level	2.893753
Respiratory problems level	1.642369
Academic performance level	2.464500
Studying workload level	1.844492
Relationship student-professor level	2.610154
Concerning with the future career	3.139055

Source: The authors (2024).

Based on the previous analysis conducted, the next subsection will describe the adjusted model, along with its characteristics and the significance of the independent variables in explaining students' stress level.

Proposed model

Initially, a multinomial logistic regression model was adjusted, considering all the 11 covariates available in the database. After this step, a preliminary analysis was carried out to determine the significance of these covariates in the model. Two of the eleven attributes did not show statistical significance, with p-values greater than 0.05. Specifically, mental health history and the level of student-teacher relationship presented p-values of 0.538 and 0.433, respectively. Given this result, a new model was adjusted considering the remaining 9 covariates, as described in Table 3.

As a consequence, the configuration of the multinomial logistic regression model adjusted for the intermediate level and high level of stress categories, while taking the low level as the reference category for students'stress, is described as follows:

$$\begin{aligned} g_1(x) &= -0.28 - 0.55(AP) + 0.36(SW) + 0.26(CF) + 0.22(CL) - 0.43(SQ) \\ &\quad +0.35(RP) + 0.12(AN) - 0.01(SF) + 0.01(DP), \end{aligned}$$

$$g_2(x) = -1.37 - 0.89(AP) + 0.61(SW) + 0.49(CF) + 0.74(CL) - 0.71(SQ) + 0.38(RP) + 0.14(AN) - 0.17(SF) + 0.08(DP). \quad (4)$$

After implementing the regression model (4), a classification analysis was carried out to evaluate its performance. Table 4 presents the values of the predictions made by the model, comparing them with the actual values in the database.

Table 4: Confusion Matrix.

	Reference		
Prevision	Low level	Intermediary level	High level
Low level	319	11	10
Intermediary level	37	334	32
High level	17	13	327

Source: from the authors (2024).

Based on the data presented in Table 4, it is possible to observe that the proposed model was successful in correctly predicting the low level category for 319 students. However, there were incorrect classifications of 11 students in the intermediate level category and 10 in the high stress level category. For the intermediate level category, the model demonstrated accuracy in predicting 334 students, however, it made 37 erroneous predictions in the low level category and 32 in the high stress level category. Finally, in relation to the high level of stress category, the model was correct in classifying 327 students, however, it made 17 incorrect predictions in the low level category and 13 in the intermediate level of stress category.

This analysis suggests that the model demonstrates considerable ability to accurately classify students according to their stress levels, highlighting the low level category as the one that achieved the best performance, recording an accuracy rate of 93.82%. In contrast, the intermediate level category presented an accuracy of 82.88%, while the high level achieved an accuracy of 91.60%. In general terms, the overall accuracy of the model in accurately classifying students according to their stress phases was 89.09%, indicating that the model has a good ability to correctly predict the categories.

Aiming to collaborate with the previous analysis, Table 5 is presented with the values of the general statistics and by category of the proposed model.

Based on the results presented in Table 5, it is observed that the model's accuracy is 0.8909, with a confidence interval of $(0.871, 0.9087)$. This means that there is 95% confidence that the true accuracy of the model lies within this range. This narrow confidence band suggests considerable precision in estimating model accuracy.

Another aspect to be highlighted is the Kappa coefficient, whose value obtained is 0.8365. This coefficient represents a significant measure of agreement between the classifications predicted by the model and those observed empirically. This result suggests robust consistency between the model's predictions and actual data, indicating its ability to make consistent and reliable predictions in relation to the different categories considered.

Furthermore, through the results presented in this table, particularly in the statistics by categories, it can be concluded that the classification model has a very satisfactory general performance for all stress levels considered. The sensitivity, specificity, and positive and negative predictive values are relatively high for each level, indicating that the model is capable of correctly identifying both positive and negative cases across all categories. Furthermore, the balanced accuracy, which is a global measure of the model's classification capacity, is also high for each stress level, which reinforces the model's effectiveness in classifying cases into all classes in a balanced way. In summary, the results suggest that the model has a good ability to discriminate between different levels of student stress.

Table 5: General statistics and by category.

General Statistics			
Accuracy	0.8909		
Confidence Interval (95% CI)	(0.871, 0.9087)		
Non-Information Rate	0.3391		
P-value (Acc > NIR)	< 2.2e-16		
Kappa	0.8365		
McNemar's Test p-valor	2.595e-05		
Nagelkerke's Pseudo-R^2	0.8243		
Statistics by Categories			
	Low level	Intermediary level	High level
Sensibility	0.8552	0.9330	0.8862
Specificity	0.9711	0.9070	0.9590
Positive Predictive Value	0.9382	0.8288	0.9160
Negative Predictive Value	0.9289	0.9656	0.9435
Prevalence	0.3391	0.3255	0.3355
Detection Rate	0.2900	0.3036	0.2973
Prevalence of Detection	0.3091	0.3664	0.3245
Balanced Accuracy	0.9132	0.9200	0.9226

Source: The authors (2024).

From the analysis of the general and specific statistics by categories of the proposed model, we continued with the evaluation of the magnitude and interpretation of its coefficients.

Magnitude and interpretation of adjusted model coefficients

Hereafter, the values of the odds ratios (OR) and confidence intervals (CI) referring to the coefficients of the proposed model are presented, considering the low stress level category as a reference for comparison.

By analyzing Table 6, it is observed that the covariates level of self-esteem, level of depression and level of headache did not present statistical significance in the intermediate level category, but exhibited a different behavior in the high level category. This distinction is evidenced by the fact that, for the intermediate level model, the odds ratio can vary within the confidence interval, including the value 1, as highlighted by Norton *et al.* (2018). All other covariates were statistically significant in both models for stress levels.

Table 6: Magnitude of the proposed model coefficients.

Intermediary level			High level		
Covariety	OR	95% CI	Covariety	OR	95% CI
Anxiety level	1.13	1.07, 1.19	Anxiety level	1.15	1.07, 1.24
Self esteem level	0.99	0.95, 1.03	Self esteem level	0.84	0.80, 0.89
Depression level	1.01	0.97, 1.06	Depression level	1.08	1.02, 1.15
Cephalea level	1.25	1.00, 1.57	Cephalea level	2.10	1.59, 2.77
Sleeping quality level	0.65	0.53, 0.80	Sleeping quality level	0.49	0.38, 0.65
Respiratory problems level	1.42	1.19, 1.70	Respiratory problems level	1.47	1.16, 1.86
Academic performance level	0.58	0.47, 0.71	Academic performance level	0.41	0.31, 0.55
Studying workload level	1.43	1.15, 1.78	Studying workload level	1.84	1.40, 2.42
Concerning with the future career level	1.29	1.03, 1.62	Concerning with the future career level	1.63	1.23, 2.17

Source: from the authors (2024).

Aiming to understand the influence of each covariate on the stress level of students during both categories, an analysis was carried out to identify both the covariates that indicate an increase in the probability of developing stress in each specific phase, as well as those that suggest a decrease in it.

Considering the aforementioned data, it is observed that each increase of one point on the scale of anxiety level, level of respiratory problems, level of study hours and level of concern about the future career is associated with an increase in the probability of a student develops stress, both at an intermediate level and at a high level, compared to a low level of stress. Specifically, it appears that each point increase on the anxiety level variable scale results in a 13% increase in the probability of a student developing an intermediate level of stress and 15% for a high level. In the case of the variable level of study hours, the probability increases are 43% and 84% for the intermediate and high levels, respectively.

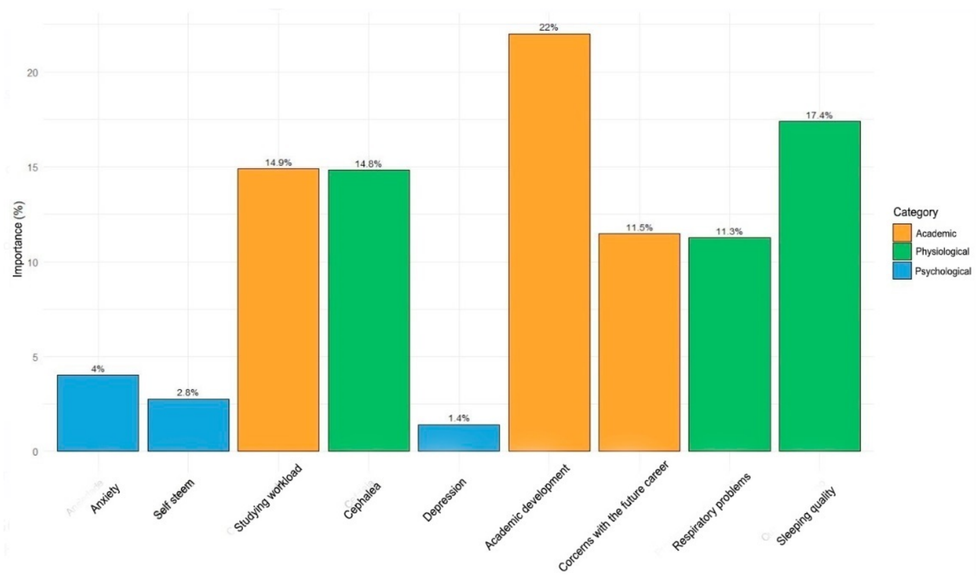
On the other hand, it is observed that each increase of one point on the scale of sleep quality level and academic performance level is associated with a decrease in the probability of a student developing stress, both for the intermediate and high levels, compared to the low level. For example, each point increase on the sleep quality level scale results in a 35% decrease in the probability of a student developing the intermediate stress level and 51% for the high level. As for the academic performance level variable, the probability decreases are even more pronounced, with 42% and 59% for the intermediate and high levels, respectively.

It is important to highlight that the depression level and headache level variables demonstrate an odds ratio greater than 1 in both stress level categories. However, as previously reported, these variables did not exhibit statistical significance at the intermediate level; only in the high level category did their values reveal significance. In this sense, it was observed that an increase of one point on the depression level scale results in an 8% increase in the probability of a student developing a high level of stress. In relation to the headache level variable, the increase in probability is more significant, reaching 110% for the high level of stress.

Significance of covariates in applying the stress level

Through an additional analysis, the relevance of the independent variables in explaining the students' stress level was examined. The results obtained in this study are presented in Figure 2 and Table 7.

Figure 2: Significance of independent variables in explaining stress level.



Source: from the authors (2024).

Table 7: Importance of variables in explaining students' stress level.

Variety	Meaningfulness (%)
Academic development	22.0
Sleeping quality	17.4
Studying workload	14.9
Cephalea	14.8
Concerning with the future career	11.5
Respiratory problems	11.3
Anxiety	4.0
Self steem	2.8
Depression	1.4

Source: from the authors (2024).

Through the results obtained in Figure 2 and Table 7, it can be inferred that the pressure for academic performance, together with the quality of sleep, the study load and the level of headache are the fundamental attributes in the context of stress among students. This fact is justified, since these variables present the highest percentages of significance in explaining the level of stress, indicating their important influence on the well-being of students. Similar results are also observed in Dávila *et al.* (2019).

Other variables such as the level of concerns about the future career and the level of respiratory problems are also relevant attributes in the students' experience of stress. Although they may not be as prominent as academic factors, they still play relevant roles, suggesting that student stress can be influenced by factors beyond the academic environment.

On the other hand, variables such as levels of anxiety, self-esteem and depression, although relevant, demonstrate a comparatively lesser importance in explaining students' stress. This suggests that, although psychological and emotional aspects are relevant, other factors may have a greater influence on the development of stress in students.

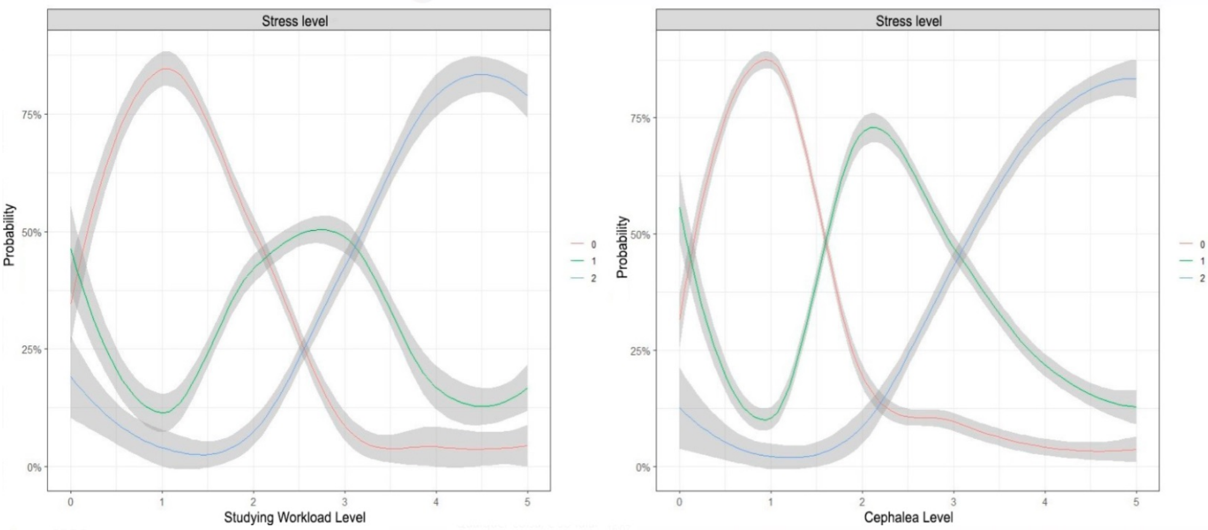
Therefore, variables of an academic nature exert the greatest influence on students' stress levels, followed by variables of a physiological nature and, lastly, those of a psychological nature. ogica. These findings suggest that, although aspects related to physical and mental/emotional health are relevant to the development of stress among students, academic variables are the most significant and have the greatest impacts on student stress. students. This result is in line with the discussions present in the works of Abacar *et al.* (2021) and Moretti and Hubner (2017), which highlight academic factors, such as the number of assessments, work and study load, as agents that increase student stress.

Considering the four most influential attributes in explaining students' stress level, an analysis was carried out to investigate the nature of this influence. The graphs contained in Figures 3 and 4 are representative illustrations of this situation. It should be noted that the curves indicated in the legend by 0, 1 and 2, respectively represent the categories low level of stress, intermediate level of stress and high level of stress.

The analysis using Figures 3 and 4 reveals distinct patterns in the probabilities of students developing stress at the three different levels. Notably, a convergence is observed in the probability values associated with the attributes of academic performance level and sleep quality level, between the different levels of stress. On the other hand, a marked disparity is evident when comparing these patterns with the variables of study load level and headache level.

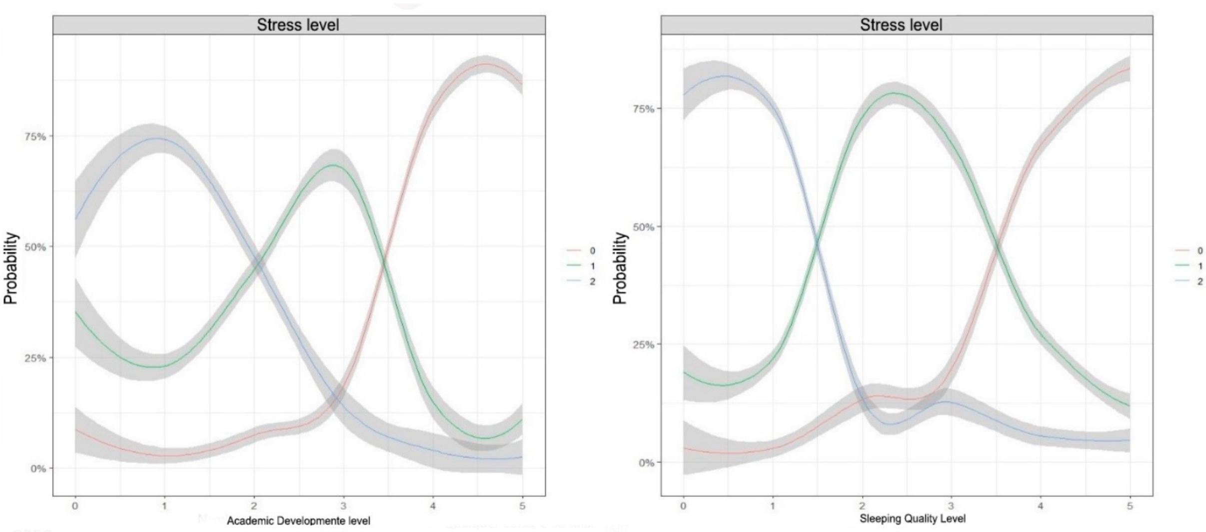
It is possible to discern, in particular, through Figure 3, patterns relating to academic performance level and sleep quality indices in relation to the probability of a student developing a low level of sleep. stress level. It is observed that as these indices increase, the probability of stress incidence at the aforementioned level also increases. Furthermore, from the level 3 value on both attributes, a substantial increase in the probability of developing a low level of

Figure 3: Behavior of the level of academic performance and the level of sleep quality in relation to the phases of the students' stress level.



Source: The authors.

Figure 4: Behavior of the study load level and the headache level in relation to the phases of the students' stress level.



Source: from the authors (2024).

stress is observed, reaching values greater than 80% for an index value equal to 5. Thus, the results suggest that the higher a student's academic performance and sleep quality, the greater the likelihood of low-stress manifestation.

It is observed, on the other hand, that for the same attributes, the lower their indices, the greater the probability of a student expressing a high level of stress. For the sleep quality level attribute, specifically, these probabilities exceed 75%. In the context of an intermediate level of stress, it appears that for intermediate indices of sleep quality and academic performance (values 2 and 3), the highest probabilities occur. On the other hand, indices close to 5 indicate lower probability values for this stress category.

In summary, based on the patterns observed in the index values of academic performance level and sleep quality in relation to the probability of a student developing one of the three levels of stress, it can be inferred that there is a significant relationship between these attributes and the manifestation of stress. Notably, the higher a student's academic performance and quality of sleep, the greater the probability of low-level stress manifestation, that is, one that does not harm health. of the subject. On the other hand, for the same attributes, the lower their indexes, the greater the chances of a student expressing a high level of stress, which can cause various pathologies in students.

Furthermore, in relation to the intermediate level of stress, it is observed that for average indices of academic performance and sleep quality, the highest probabilities occur. However, indices with values close to 5 indicate lower probabilities for this stress category. These patterns suggest the importance of the balance between academic performance and sleep quality in regulating stress in different phases, highlighting that good levels of academic performance and sleep quality directly interfere with the state student stress.

Regarding to Figure 4, there are distinct patterns in relation to the values of study load and headache levels and their associations with the different stress phases. For levels with a value of 1 of these attributes, it is observed that the chances of a student developing a low level of stress are significantly high, exceeding 75% in both attributes. However, as the values of the levels of these variables increase, the probabilities of developing stress in this category tend to decrease.

However, this behavior changes when the probability of developing a high level of stress is analyzed. From the level 1 value of these attributes, there is a significant increase in the probability of developing a high level of stress. In other words, the higher the values of the study load and headache indices, the greater the chances of a student developing a high level of stress. As for the intermediate level of stress, it appears that for median indices of these attributes (values 2 and 3), their highest probabilities occur. Specifically, for the study load level, this probability is approximately 50%, while for the headache level it is approximately 75%.

In summary, this stage of the investigation revealed distinct patterns in the probabilities of developing stress at different levels, especially for low and high levels of stress. While better academic performance and sleep quality seem to be associated with a greater probability of developing low levels of stress, lower levels of these attributes are correlated with a greater chance of developing high levels of stress.

Concluding Remarks

Through this study, it was possible to verify a high prevalence of stress among Nepalese students, so that approximately one third of them face levels that affect their daily lives. Thus, seeking to identify the factors and their influence on stress levels among Nepali students, a multinomial logistic regression model was implemented. The constructed model proved to be effective in classifying students according to their stress levels, presenting an overall accuracy of 89.09%.

The results highlighted the significant influence of factors such as anxiety, respiratory problems, study hours and concern about a future career in increasing stress, while sleep quality and academic performance were associated with a decrease dog in the probability of stress. In particular, the attributes, academic performance, quality of sleep, study load and headache can be highlighted as the most significant in explaining students' stress level. These results are important in supporting policies and programs that aim to promote the well-being of students and deal with academic pressures.

In addition to the aforementioned, this investigation highlights the need for interventions aimed at students' mental health, including accessible counseling services and programs to promote mental health. emotional health. The results also highlight the importance of addressing not only academic factors, but also the psychological and physical aspects that contribute to student stress. Investing in self-care strategies, education on stress management and creating an emotionally supportive environment can help mitigate the negative effects of stress on students' lives and promote a study environment healthier and more productive.

References

ABACAR, M.; ALIANTE, G.; ANTÓNIO, J. F. *Stress e estratégias de coping em estudantes universitários.* Aletheia, v.54, n.2, 2021.

ABREU, K. L. D.; STOLL, I.; RAMOS, L. S., BAUMGARDT, R. A.;KRISTENSEN, C. H. *Estresse ocupacional e Síndrome de Burnout no exercício profissional da psicologia.* Psicologia: ciência e profissão, v.22, p.22-29, 2002.

BORINE, R. C. C.; WANDERLEY, K. S.; BASSIT, D. P. *Relação entre a qualidade de vida e o estresse em acadêmicos da área da saúde.* Estudos Interdisciplinares em Psicologia, v.6 n.1, p.100-118, 2015.

BUBLITZ, S.; GUIDO, L. D. A.; LOPES, L. F. D.; FREITAS, E. D. O. *Association between nursing students'academic and sociodemographic characteristics and stress.* Texto & Contexto-Enfermagem, vol.25, 2016.

DAS MANGAS, E. J. *Introdução à regressão categorial aplicação a dados de escolas angolanas, com utilização do SSPS.* (Tese de doutorado). Universidade da Beira Interior, Portugal, p. 64, 2019.

DÁVILA, O. S.; GOTARATE, F. B.; SOLANO, O. B. *Estresse acadêmico em estudantes universitários usando análise multivariada.* Sigmae, v.8, n.2, p. 348-358, 2019.

DE LIMA, I. M. B.; DE OLIVEIRA, A. E. C.; DE ANDRADE, J. M.; COÊLHO, H. F. C.; LIMA, K. S. *Modelo de decisão sobre o uso de preservativos: uma regressão logística multinomial.* Revista Tempus-Actas de Saúde Coletiva. 10(2), p.67-80, 2016.

ELTINK, C. F.; NUNES, C. W. B. *Concepções sobre estresse segundo alunos do ensino médio de uma cidade de pequeno porte.* Revista Família, Ciclos de Vida e Saúde no Contexto Social, vol. 1, p. 549-561, 2020.

FÁVERO, L. P. L.; BELFIORE, P. P.; SILVA, F. L. D.; CHAN, B. L. *Análise de dados: modelagem multivariada para tomada de decisões*, Rio de Janeiro: Elsevier, 2009.

HOSMER, D. W., LEMESHOW, S., STURDIVANT, R. X. *Applied logistic regression.* 3.ed. Hoboken (NJ): Wiley & Sons, Inc., 2013.

JOHNSON, R.; WICHERN, D. W. *Applied Multivariate Statistical Analysis.* New Jersey: Prentice Hall International, Inc. 1988.

JOYOTSNA, C.; AMUDHA, J. *Eye gaze as an indicator for stress level analysis in students.* In 2018 International conference on advances in computing, communications and informatics (ICACCI), p. 1588-1593, IEEE, 2018.

JUST, A. P., ENUMO, S. R. F. *Problemas emocionais e de comportamento na adolescência: o papel do estresse.* Boletim Academia Paulista de Psicologia, vol.35, n^o 89, 2015.

MARTINS, R. M. B.; BOMFIM, F. N. G.; GOUVEIA, L. Q.; COSTA SPONTAN, O.; DE MELO, C. M. I.; BEZERRA, L. A.; PIMENTEL, D. M. M. *Estresse em alunos de preparatórios para vestibular.* Brazilian Journal of Health Review, 4(3), p. 10639-10651, 2021.

MORETTI, F. A.; HUBNER, M. M. C. *O estresse e a máquina de moer alunos do ensino superior: Vamos repensar nossa política educacional?* Revista de Psicopedagogia, 34 (105), 258-67, 2017.

NORTON, E. C.; DOWD, B. E.; MACIEJEWSKI, M. L. *Odds ratios?current best practice and use.* Jama, v. 320, n. 1, p. 84-85, 2018.

PARIAT, L.; RYNJAH, A.; JOPLIN, M.; KHARJANA, M. G. *Stress levels of college students: Interrelationship between stressors and coping strategies.* Journal of Humanities and Social Science, 19(8), p. 40-46, 2014.

PITT, A.; OPRESCU, F.; TAPIA, G.; GRAY, M. *An exploratory study of students? weekly stress levels and sources of stress during the semester.* Active Learning in Higher Education, 19(1), p. 61-75, 2018.

SANTOS, F. S.; MAIA, C. R. C.; FAEDO, F. C.; GOMES, G. P. C.; NUNES, M. E.; OLIVEIRA, M. V. M. D. *Estresse em estudantes de cursos preparatórios e de graduação em medicina.* Revista Brasileira de Educação Médica, 41(2), p. 194-200, 2017.

SOFRO, A.; KHIKMAH, K. N.; ARIYANTO, D.; FUAD, Y.; RAHADJENG, B.; ASTUTI, Y. P. *Handling Overdispersion Problems in Multinomial Logistic Regression (Study Case in Stress Level Data).* Wseas, vol.3, p.78-83, 2023.