

<https://doi.org/10.23913/ride.v15i30.2246>

Scientific articles

Identificación de Causas del Estrés Académico en Estudiantes de la UPMH mediante un Modelo de Regresión Logística

Identification of Causes of Academic Stress in UPMH Students Using a Logistic Regression Model

Identificação de causas de estresse acadêmico em alunos da UPMH por meio de modelo de regressão logística

Glendamira Serrano Franco

Universidad Politécnica Metropolitanan

gfranco@upmh.edu.mx

<https://orcid.org/0000-0003-3176-3433>

Víctor Manuel Zamudio García

Universidad Politécnica Metropolitanan

vzamudio@upmh.edu.mx

<https://orcid.org/0000-0002-4660-8025>

Resumen

Esta investigación se centra en la Universidad Politécnica Metropolitana de Hidalgo (UPMH), donde se ha identificado un aumento en los niveles de estrés entre los estudiantes, lo que ha llevado a un incremento en la reprobación y la deserción. La pregunta de investigación se enfoca en cómo identificar, mediante un modelo de regresión logística, las principales causas de estrés en los estudiantes de cuatro programas educativos de la UPMH. El objetivo es determinar estas causas utilizando dicho modelo. El diseño de la investigación es no experimental, transversal, descriptivo y de enfoque cuantitativo. Los datos se recopilaron en un único momento a través de un cuestionario administrado mediante un muestreo no probabilístico, con una muestra de quinientos cuarenta y dos estudiantes. Según los resultados, más del sesenta por ciento de los estudiantes manifestaron sentir estrés. La integración de los resultados en el modelo de regresión logística permitió identificar las principales causas del estrés, siendo la carga académica

durante cada cuatrimestre y la responsabilidad de realizar otras actividades, como trabajar y/o participar en actividades extracurriculares, las más relevantes.

Palabras clave: carga académica, deserción, estrés, logística, modelo de regresión, reprobación.

Abstract

This research focuses on the Universidad Politécnica Metropolitana de Hidalgo (UPMH), where an increase in stress levels among students has been identified, leading to higher rates of failure and dropout. The research question focuses on how to identify, through a logistic regression model, the main causes of stress in students from four educational programs at UPMH. The objective is to determine these causes using the aforementioned model. The research design is non-experimental, cross-sectional, descriptive, and quantitative. Data were collected at a single point in time through a questionnaire administered using non-probabilistic sampling, with a sample of five hundred forty-two students. According to the results, more than sixty percent of the students reported feeling stressed. The integration of the results into the logistic regression model allowed the identification of the main causes of stress, with the most relevant being the academic workload during each semester and the responsibility of performing other activities, such as working and/or participating in extracurricular activities.

Keywords: academic workload, dropout, stress, logistics, regression model, failure.

Resumo

Esta investigação centra-se na Universidade Politécnica Metropolitana de Hidalgo (UPMH), onde se observou um aumento nos níveis de stress entre os estudantes, resultando em um maior índice de reprovação e evasão. A questão de investigação centra-se em como identificar, por meio de um modelo de regressão logística, as principais causas do stress em estudantes de quatro programas educativos da UPMH. O objetivo é determinar essas causas utilizando o referido modelo. O desenho da pesquisa é não experimental, transversal, descritivo e com abordagem quantitativa. Os dados foram coletados em um único momento por meio de um questionário aplicado por amostragem não probabilística, com uma amostra de quinhentos e quarenta e dois estudantes. Segundo os resultados, mais de sessenta por cento dos estudantes relataram sentir stress. A integração dos resultados no modelo de regressão logística identificou as principais causas do stress, destacando-se a



carga acadêmica em cada semestre e a responsabilidade de trabalhar e/ou participar de atividades extracurriculares.

Palavras-chave: carga acadêmica, evasão, stress, logística, modelo de regressão, reprovação.

Reception Date: July 2024

Acceptance Date: January 2025

Introduction

Stress has been a constant in human experience, faced by adapting to difficult situations in order to survive. According to Cote and García (2016), stress is part of everyday life and constitutes a normal and adaptive reaction. Porras et al. (2024) describe the term as an adaptation into Spanish of "stress", introduced by Hans Selye in 1926, who defined it as the general response of the organism to stressful stimuli.

This phenomenon manifests itself not only in exceptional circumstances, but also in everyday aspects such as workload, academic pressure, financial concerns, and technological demands. Espinosa et al. (2020) mention that high levels of stress in university students represent obstacles to academic performance. Calatayud et al. (2021) emphasize that stress, along with loneliness and interpersonal mistrust, is associated with depression. Finally, Barrón and Armenta (2021) define academic stress as a state that arises when activities exceed the student's resources or capabilities, which affects their thinking, learning, and productivity.

An Exploration of Research References

Over time, the cumulative effects of stress can lead to health problems ranging from the common cold to serious illness. To support this research, the following references were considered:

Ramírez et al. (2022) integrated data mining techniques, such as hierarchical cluster analysis and logistic regression, to characterize students at the Autonomous University of the State of Hidalgo. Through a survey, they collected sociodemographic, economic, technical, and mental health data, which allowed them to identify patterns that influence academic performance. The results revealed the existence of two distinct clusters and a list of factors that positively and negatively impact academic performance.

For their part, Peralta et al. (2022) applied an unsupervised K-Means algorithm to analyze depression, anxiety and stress using the DASS-21 Scale in university students from

Apurímac-Peru. They found that the level of depression (60%) was more prevalent than that of anxiety and stress, with a higher incidence at the moderate level.

Problematic

This research focuses on the Metropolitan Polytechnic University of Hidalgo, where high levels of stress have been detected in students of various educational programs. This situation has negatively impacted academic performance, increasing failure and dropout rates. Given this problem, the following question arises:

What are the causes of high levels of stress in students of four educational programs at the Metropolitan Polytechnic University of Hidalgo, and how can they be identified using a logistic regression model?

Aim

To identify, using a logistic regression model, the causes that generate high levels of stress in students of four educational programs at the Metropolitan Polytechnic University of Hidalgo.

Hypothesis

The integration of a logistic regression model will allow us to identify the causes of high levels of stress in students of four educational programs at the Metropolitan Polytechnic University of Hidalgo.

Materials and methods

Type of study

The research adopts a quantitative approach, defined by Sánchez (2019) as a method that addresses measurable phenomena using statistical techniques to analyze data. This approach seeks to describe, explain, predict and control causes, as well as anticipate their occurrence. According to Alan and Cortez (2018), this type of analysis allows for generating discussions and publications based on statistical results.

The selected design is a descriptive non-experimental design, which, according to Hernández et al. (2017), is valuable for accurately presenting aspects or dimensions of a phenomenon, context, or community. This design requires the researcher to define what will be measured and who the data will be collected from.

Information gathering technique

According to Castellano et al. (2024), data collection includes procedures that allow obtaining relevant information to answer the research question. Hernández and Duana (2020) describe this process as systematic measurement to acquire scientific knowledge.

Instrument for data collection

A questionnaire was used as the main instrument, following the recommendations of Medina et al. (2023). This questionnaire included questions about academic workload, study hours, extracurricular activities, employment to finance studies, and stress symptoms. The aim was to identify the main causes of stress among students. The data collected were analyzed using a logistic regression model to determine the underlying causes of high stress levels.

Sample

The study covered four Educational Programs: Information Technology Engineering, Animation and Visual Effects Engineering, Bachelor's Degree in Business Management and Bachelor's Degree in International Trade and Customs, with a total sample of 542 students. The sampling was non-probabilistic and by convenience, following the Arrogante criterion (2021).

- Information Technology Engineering: 82 students selected from an enrollment of 118.
- Business Management: 125 students selected from 152.
- Animation and Visual Effects Engineering: 156 students selected from 183.
- International Trade and Customs: 179 students selected from 206

Data Analysis and Processing

The data were analyzed using a Logistic Regression Model, defined by Martínez and Pérez (2024) as a statistical technique to verify hypotheses and explore causal relationships between categorical and independent dependent variables.

Development of the Logistic Regression Model

According to Ossa and Jaramillo (2021), logistic regression is a method for binary classification problems. In this study, it was used to categorize stress levels into “high” and “low”. The model was implemented with machine learning libraries, such as scikit-learn, in the Google Colaboratory tool.

Steps followed in the development of the model:

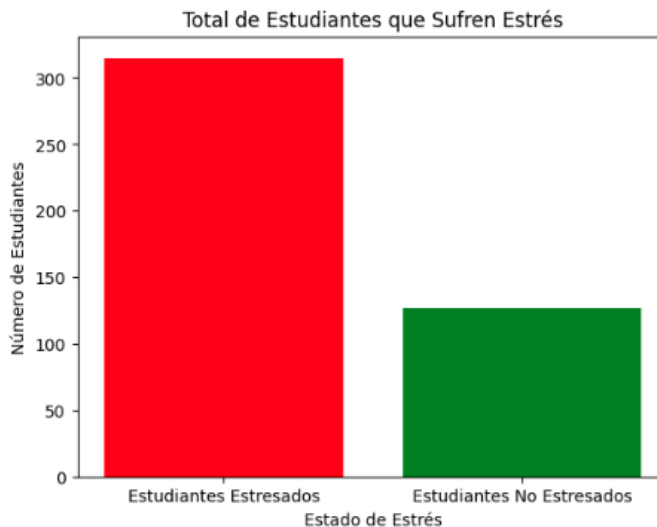
1. Definition of the theoretical framework: The literature on academic stress and associated variables was reviewed.
2. Selection of variables: Factors such as academic load, performance, extracurricular activities and physical or emotional symptoms were identified.
3. Questionnaire design: A clear and relevant questionnaire was developed to measure these variables.
4. Application of the questionnaire: Administered to 542 selected students.
5. Data preprocessing: Included coding, handling of missing data, and normalization.
6. Classification model development: Implementation of the algorithm with adjustments to optimize performance.
7. Validation and evaluation: Metrics such as accuracy, sensitivity and specificity were applied.
8. Interpretation of results: Identification of significant patterns that contribute to student stress.

Results

The results obtained using the logistic regression model are detailed below. Figure 1 presents the analysis of the model, including the accuracy obtained in the training and test data sets. These results allowed us to identify the total number of students with high and low levels of stress, considering the associated variables. These findings offer a detailed perspective of the relationship between the variables analyzed and the stress levels in the students .

Figure 1. Students with high and low stress levels

Precisión en el conjunto de entrenamiento: 0.7932011331444759
 Precisión en el conjunto de prueba: 0.797752808988764
 Total de estudiantes estresados: 315
 Total de estudiantes no estresados: 127



Source: Own elaboration

The results were categorized according to the Educational Program, with the aim of identifying the students with the highest stress index based on their area of study. Figure 2 shows a classification that groups the students by age, pressure experienced and level of stress detected, considering the variables analyzed.

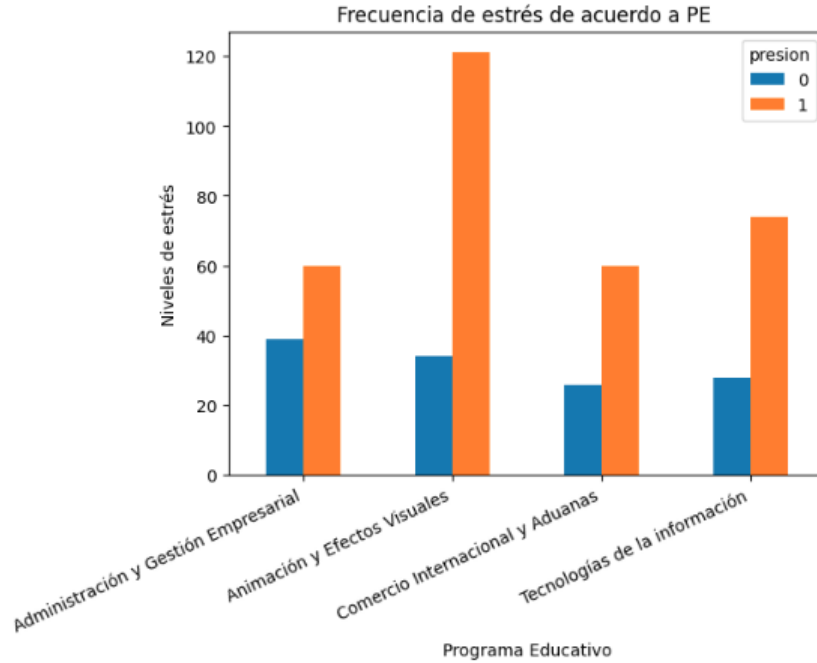
Figure 2. Stress Level by Educational Program

	edad	presion	nivel_estres
PE			
Administración y Gestión Empresarial	19.696970	0.606061	6.292929
Animación y Efectos Visuales	20.245161	0.780645	6.367742
Comercio Internacional y Aduanas	19.581395	0.697674	6.139535
Tecnologías de la información	19.764706	0.725490	6.401961

Source: Own elaboration

Figure 3 classifies students according to their educational program and the level of stress identified by the model. The blue bars represent students with a low level of stress, while the orange bars indicate those with a high level of stress. The results highlight that students in the Animation and Visual Effects Educational Program present the highest levels of stress .

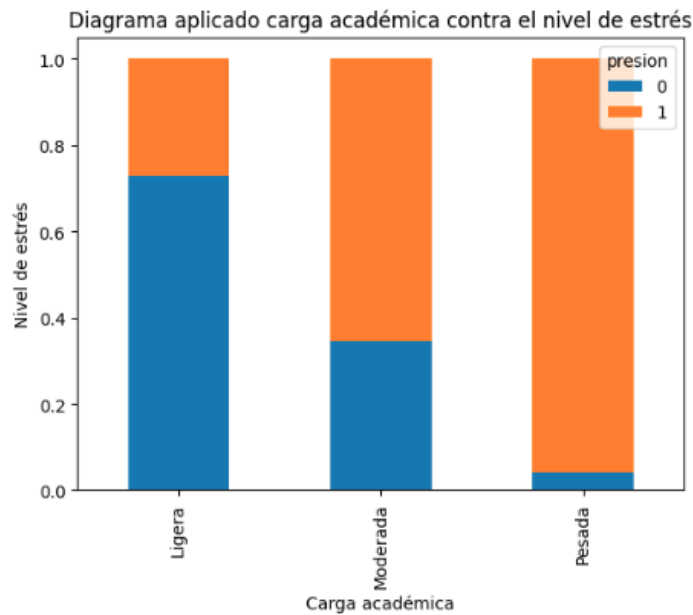
Figure 3. Classification of stress levels by Educational Program



Source: Own elaboration

Among the main causes of stress, academic workload was identified as a significant factor. Figure 4 shows the relationship between this variable and the increase in stress levels in students.

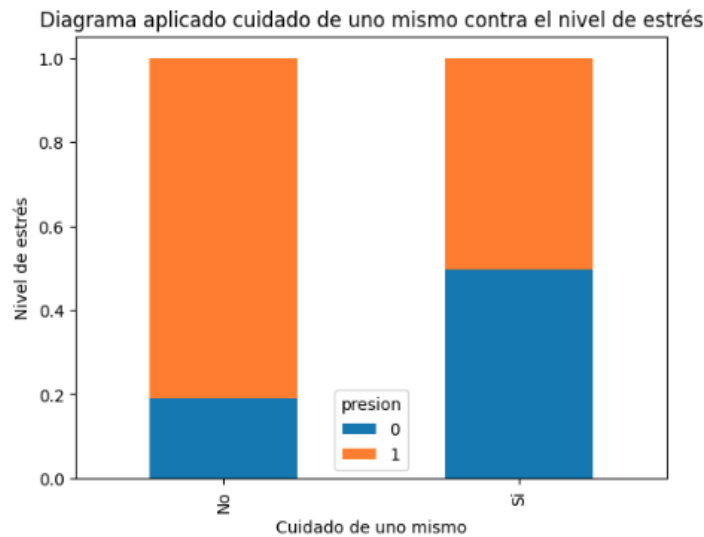
Figure 4. Relationship between academic load and stress levels



Source: Own elaboration

Finally, the variable of time availability for self-care was analyzed. The results, presented in Figure 5, indicate that high levels of stress make it difficult for students to devote sufficient time to caring for their physical and psychological health.

Figure 5. Relationship between time for self-care and stress levels



Source: Own elaboration

Summary of Results

- **Academic Load:** Academic load was identified as the main cause of increased stress among students, highlighting the need to strategically redistribute academic tasks and responsibilities.
- **Other Sources of Stress:** Extracurricular activities and the need to work are also highlighted as significant factors, as they limit the time available for students' academic and personal activities.
- **Variation across Educational Programs:** Significant variation in stress levels was observed across educational programs. Students in the Animation and Visual Effects Program had the highest levels of stress compared to the others.

The application of the logistic regression model was effective in identifying the underlying causes of high levels of stress among students. These results emphasize the importance of implementing specific strategies to mitigate stress, such as the redistribution of academic loads and the design of activities that promote balance between academic, work and personal responsibilities.

Discussion

The initial problem of this research was that high levels of stress among students lead to poor academic performance. To answer this question, the causes of stress in students from four educational programs at the Metropolitan Polytechnic University of Hidalgo (UPMH) were investigated by applying a logistic regression model.

The analysis of the results allows us to conclude that an excessive workload and academic responsibilities is one of the main causes of stress in students. This, in turn, generates negative consequences such as limited learning, extreme fatigue and alterations in sleep patterns, which directly affects their academic performance. These findings coincide with the research of Barrón and Armenta (2021), who describe that academic stress arises when demands exceed the student's resources or capabilities, resulting in a decrease in productivity and learning.

The integration of the logistic regression model in this study allowed to effectively identify the main causes of academic stress. Likewise, the use of advanced technological tools, such as artificial intelligence, was shown to be useful in predicting the physiological effects related to stress. This approach is supported by the findings of Mora and Martell (2021), who employed a multilayer perceptron artificial neural network to model and predict the physiological effects associated with academic stress.

Conclusions

In conclusion, the integration of a logistic regression model allowed us to effectively identify the causes that generate high levels of stress in students. The results support the hypothesis raised in this research, since the main causes of stress were identified, which manifest themselves in physical and emotional symptoms that negatively affect academic performance.

The analysis of the model revealed that, out of a total of five hundred and forty-two students, one hundred and seventy-seven had low levels of stress, while three hundred and sixty-five had high levels. A significant number of the latter belonged to the Animation and Visual Effects Educational Program. In addition, it was identified that academic workload and extracurricular activities are the main causes of stress, especially when students do not have adequate time management.

Future Lines of Research

As part of the strategies to mitigate student stress, a multisensory space is being developed designed to stimulate the five senses and reduce stress levels among students. This project, currently in the process of implementation, seeks to benefit the student community of the Metropolitan Polytechnic University of Hidalgo.

References

- Arrogante, O. (2021) Técnicas de muestreo y cálculo del tamaño muestral: Cómo y cuántos participantes debo seleccionar para mi investigación. *Enfermería Intensiva*, 33(1), 44-47. doi: <https://doi.org/10.1016/j.enfi.2021.03.004>
- Barrón, M. y Armenta, L. (2021) Efectos del Estrés en el Rendimiento Académico de los Estudiantes de la Licenciatura de Psicología. *Revista de Investigación Académica sin Frontera*. 1(36). doi: <https://doi.org/10.46589/rdiasf.vi36.388>
- Calatayud, A., Apaza, E., Huaquisto, E., Belizario, G. y Inquilla, J. (2021). Estrés como factor de riesgo en el rendimiento académico en el estudiantado universitario (Puno, Perú). *Revista Educación*. 46(2), 2-3. doi: <https://doi.org/10.15517/revedu.v46i2.47551>
- Castellano, M., Coúrbenas, D., Leal, R. y Pérez, M. (2024) Técnicas e instrumentos para la recolección de datos de la investigación. *Universidad Pedagógica Experimental Libertador*. Sitio web: <https://acortar.link/ORm5C0>
- Cote, L. y García, A. (2016) Estrés como factor limitante en el proceso de toma de decisiones: una revisión desde las diferencias de género. *Avances en Psicología Latinoamericana*. 34(1), 19-28. doi: <https://dx.doi.org/10.12804/apl34.1.2016.02>
- Espinosa, J., Hernández, J., Rodríguez, J., Chacín, M. y Bermúdez, V. (2020) Influencia del estrés sobre el rendimiento académico. *Revista Archivos Venezolanos de Farmacología y Terapéutica*. 39(1), 64-65. doi: <https://doi.org/10.5281/zenodo.4065032>
- Hernández, R., Méndez, S., Mendoza, C. y Cuevas, A. (2017) Fundamentos de Investigación. Editorial: McGraw-Hill. México. Sitio web: <http://librodigital.sangregorio.edu.ec/librosusgp/B0035.pdf>
- Hernández, S. y Duana, D. (2020) Técnicas e instrumentos de recolección de datos. *Boletín Científico de las Ciencias Económico Administrativas del ICEA*. 9(17). Sitio web: <https://repository.uaeh.edu.mx/revistas/index.php/icea/article/view/6019/7678>

- Martínez, J. y Pérez, P. (2024) Regresión Logística. *Medicina de familia*. 50(1). doi: <https://doi.org/10.1016/j.semerg.2023.102086>
- Medina, M., Rojas, R., Bustamante, W., Loiza, R., Martel, C. y Castollo, R. (2023) Metodología de la investigación: Técnicas e instrumentos de investigación. *Instituto Universitario de Innovación Ciencia y Tecnología*. Inudi, Perú, S.A.C. doi: <https://doi.org/10.35622/inudi.b.080>
- Mora, J. y Martell, J. (2021) Predicción de efectos fisiológicos causados por el estrés académico mediante redes neuronales artificiales. *Revista Iberoamericana de Psicología*. 1(14). doi: <https://doi.org/10.33881/2027-1786.rip.14303>
- Ossa, W. y Jaramillo, Verónica (2021) *Machine Learning para la estimación del riesgo de crédito en una cartera de consumo*. (Tesis de Magíster en Administración Financiera) Universidad EAFIT. Medellín. Sitio web: <https://repository.eafit.edu.co/server/api/core/bitstreams/324d9a52-33ab-4255-bfaf-132f206c753e/content>
- Peralta, M., Merma, J., Chávez, E., Soto, C. Jiménez, W. (2022) *Aplicación del Algoritmo K-Means en estudiantes universitarios del área de sistemas e informática para caracterizar la salud mental, durante el aislamiento en COVID-19*. Memorias de la Vigésima Primera Conferencia Iberoamericana en Sistemas, Cibernética e Informática. doi: <https://doi.org/10.54808/CISCI2022.01.92>
- Porras, F., Guzmán, M., Barragán, L., Quintanar, M., Linares, O. y Garza, L. (2024). Estrés laboral, burnout y factores asociados a su aparición en técnicos en urgencias médicas. *Psicología y Salud*. 34(1), 93-101. doi: <https://doi.org/10.25009/pys.v34i1.2847>
- Ramírez, L., Delgado, E. y Montúfar, M. A. (2022). Aplicación de técnicas de minería de datos para la caracterización de estudiantes bajo el efecto de la COVID-19. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*. 10(2). doi: <https://doi.org/10.29057/icbi.v10iEspecial2.8669>
- Sánchez, F. (2019) Fundamentos Epistémicos de la Investigación Cualitativa y Cuantitativa: Consensos y Disensos. *Revista Digital de Investigación en Docencia Universitaria*. 13(1), 102-122. doi: <https://doi.org/10.19083/ridu.2019.644>

Contribution Role	Author(s)
Conceptualization	Glendamira Serrano Franco
Methodology	Glendamira Serrano Franco
Software	Glendamira Serrano Franco
Validation	Victor Manuel Zamudio Garcia
Formal Analysis	Glendamira Serrano Franco
Investigation	Glendamira Serrano Franco
Resources	Victor Manuel Zamudio Garcia
Data curation	Glendamira Serrano Franco
Writing - Preparing the original draft	Glendamira Serrano Franco
Writing - Review and editing	Victor Manuel Zamudio Garcia
Display	Glendamira Serrano Franco
Supervision	Glendamira Serrano Franco
Project Management	Victor Manuel Zamudio Garcia
Acquisition of funds	Glendamira Serrano Franco