Factors Influencing Obesity Levels in Mexico, Peru, and Colombia

Chen Yang *

Rutgers, the State University of New Jersey, 57 US Highway 1, New Brunswick, NJ 08901, United States

* Corresponding Author Email: cy343@scarletmail.rutgers.edu

Abstract. The prevalence of obesity has reached alarming proportions on a global scale, exhibiting a marked upward trajectory in developed countries and developing countries. The objective of this research paper is to analyze the principal factors influencing obesity in individuals between the ages of 14 and 61 in Mexico, Peru, and Colombia, utilizing an obesity risk dataset. A variety of socioeconomic, behavioral, and genetic factors were considered, including variables such as age, caloric intake, physical activity, and family history of overweight. Exploratory data analysis (EDA) was conducted using visualization techniques to elucidate the interrelationships between these variables and obesity levels. Subsequently, a logistic regression model was constructed to estimate the probability of different obesity levels based on the predictor variables. The model was evaluated using accuracy, a confusion matrix, and a receiver operating characteristic (ROC) curve analysis. The results demonstrated that the model exhibited a robust capacity to differentiate between obesity levels but encountered notable challenges in distinguishing between adjacent categories (e.g., "Overweight Level I" and "Overweight Level II"). This study offers invaluable insights into the complex multifactorial nature of obesity and provides guidance for public health interventions in Latin American countries to address the escalating obesity epidemic.

Keywords: Factors influencing obesity levels, Mexico, Peru, Colombia.

1. Introduction

The prevalence of obesity has significantly increased in developed countries and developing countries, representing a significant challenge to global public health. Given the association between obesity and a lot of health complications, such as cardiovascular disease, diabetes, and reduced life expectancy, it is imperative to gain insight into the underlying factors contributing to this condition. In Latin America, obesity prevalence has increased significantly in countries such as Mexico, Peru, and Colombia. This is the result of a complex interplay between socioeconomic, behavioral, environmental, and genetic factors.

Socioeconomic factors, including income, education, and gender, exert a significant influence on health outcomes and thus represent a crucial area of focus in the study of obesity. For example, higher income levels often mean greater access to nutritious food and healthcare, while lower income levels mean higher obesity rates due to limited healthy food choices. In addition, educational attainment significantly influences dietary choices and lifestyle, with higher-educated people generally making healthier decisions. Behavioral factors such as physical inactivity and unhealthy eating habits are also directly linked to obesity, particularly in urban areas where fast food is prevalent [1].

The research team led by Dr. Lee conducted a thorough analysis of the various factors contributing to adolescent obesity [2]. The analysis focused on three key areas: demographic and socioeconomic characteristics, behaviors, and education. The study underscored the pivotal role of demographic factors, including age, gender, and place of residence, in shaping adolescent obesity rates. Young people living in urban areas are more likely to be obese than their rural peers, a phenomenon that can be attributed to differences in lifestyle and eating habits. Socioeconomic factors, which including income and parent’s education level, also exert a significant influence. A higher socioeconomic status is often associated with greater ease of access to healthy foods and recreational facilities. In addition, the study highlights the key role played by health behaviors such as physical activity and eating habits in determining the prevalence of obesity. Regular exercise and a balanced diet are associated with a reduced likelihood of obesity in adolescents.

In addition to their primary research, Fan and Hossain also offered further insights into the intricate nuances of obesity. Fan's team studied the impact of abdominal obesity on cognitive function, and the results showed that obesity may have a significant impact on physical health [3]. The latter study identified genetic factors that influence obesity and highlighted the key role played by G-protein coupled receptors in regulating body weight and metabolism in response to a high-fat diet [4].

Using the Obesity Risk Data Set, this study investigated the causes of obesity in Mexico, Peru, and Colombia from adolescence to middle age. The objective of this research is to identify the principal determinants of obesity, considering variables such as age, diet, physical activity, and socioeconomic status. This comprehensive analysis aims to identify patterns and associations to inform the development of public health strategies and interventions to address the specific needs of these populations.

The value of this study lies in its potential to fill a gap in the existing literature by providing a comprehensive picture of the factors contributing to obesity in Latin America. The goal is to provide some insights and try to provide a basis for the development of effective public health policies and programs to promote healthier lifestyles by combating obesity in these countries. This will be achieved by integrating data analysis with socio-behavioral insights.

2. Methodology

2.1. Data Source

The dataset utilized in this research was procured from the publicly accessible "Obesity Levels Data Set," which encompasses demographic, behavioral, and physical measurements of individuals between teenagers and middle-aged from Latin America [5]. The dataset comprises a range of features on dietary habits, physical activity levels, and personal information, thereby facilitating an investigation into the factors contributing to obesity. It is noteworthy that 77% of the dataset was generated by using the Weka and SMOTE filters, to address issues related to class imbalance. The remaining 23% was obtained directly from the website. The combination of real-world and synthetically balanced data ensures that the dataset encompasses a comprehensive range of obesity levels and minimizes potential bias, enhancing its reliability for classification tasks.

2.2. Data Preprocessing

The process of data preprocessing comprises a series of essential stages designed to guarantee the integrity and uniformity of the data set. Initially, individual entries between the ages of 14 and 61 were excluded to maintain a focus on the target population. Missing data may be filled by deleting incomplete cases or by employing the mean or median to fill in missing values. This approach helps to retain as much data as possible while ensuring statistical validity. Furthermore, the paper includes some categorical variables such as gender, family's overweight history, smoking habits (SMOKE), consuming high-calorie foods frequently (FAVC), and mode of transport (MTRANS), which were transformed into factor variables to enhance the efficiency of statistical modeling. Continuous variables, including age, height, weight, and calorie intake, were standardized by the specifications of the model under consideration.

2.3. Statistical Analysis

The methodology employed in this study comprises two principal phases: exploratory data analysis (EDA) and statistical modelling. Both were indispensable for the elucidation of the factors that exert an influence on obesity levels. The exploratory data analysis (EDA) was employed as a preliminary step to provide an overview of the data set and identify patterns and associations between variables. The visualizations, created using the "ggplot2" and "corrplot" libraries in R, facilitated the exploration of the distribution of variables and their relationships with obesity levels. Bar charts demonstrated the prevalence of obesity across diverse populations, including gender and family history of overweight. Boxplots, on the other hand, delineated the variations in caloric intake, physical activity frequency, and time spent using technology devices among individuals with varying obesity levels. Scatter plots are used to study the relationship between variables, such as age and weight, and their effect on obesity. Moreover, correlation matrices were constructed for continuous variables, including age, weight, height, caloric intake, and physical activity, to elucidate significant relationships among these factors. Subsequently, a logistic regression model was constructed to ascertain the principal predictors of obesity. Logistic regression was selected as the most appropriate method for analyzing categorical outcome variables, such as obesity levels. The objective of the model was to evaluate the different obesity categories' probability based on predictor variables. For example, age, caloric intake, physical activity, and various behavioral and socioeconomic factors. The logistic regression model is defined as:

$$log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{1}$$

In the model, P (Y = 1) represents the individual's probability classified into the obesity category. The independent variables (X1, X2, …, Xn) are the predictor variables. It also measures the strength and direction of the relationship between the predictor variables and the obesity outcome by using (β0, β1, …, βn). The choice of logistic regression was based on its interpretability, whereby the coefficients can be understood in terms of odds ratios, thus providing a clear indication of the contribution of each predictor variable to the likelihood of obesity. Moreover, it is applicable to both continuous and categorical variables, rendering it an appropriate choice for this data set. It is renowned for its robustness, particularly in instances where the relationships between variables may not be strictly linear. The model was then applied to the data set, and the resulting output was assessed for statistical significance and the strength of the relationships between the variables of predictor and the outcome of obesity. To assess the efficacy of the logistic regression model, the research team employed a range of performance indicators. The accuracy was calculated to determine the proportion of all instances that were correctly classified. The performance of the model was described using a confusion matrix. Moreover, the ROC curve was employed to graphically illustrate the compromise between sensitivity and specificity, thus furnishing a grasp of the model's efficacy in differentiating between disparate obesity categories.

By integrating both electronic data analysis (EDA) and logistic regression modeling, this study provides a thorough analysis of the factors. These factors influence obesity levels in Latin American populations. It is anticipated that the results will inform the formulation of public health strategies designed to mitigate the rising prevalence of obesity in Mexico, Peru, and Colombia.

3. Results and Discussions

Through exploratory data analysis (EDA), the factors influencing obesity levels can be identified by using charts.

As illustrated in Figure 1, the correlation matrix elucidates the interrelationships between the continuous variables. There is a strong positive correlation between height and weight. The frequency of vegetable consumption (FCVC) is moderately negatively correlated with weight, indicating that consuming more vegetables is associated with lower weight, which is consistent with the understanding that a vegetable-rich diet helps with weight management. The frequency of physical activity (FAF) is also slightly negatively correlated with weight, which suggests that regular physical activity helps with weight control. Time spent using technology (TUE) was weakly positively associated with weight, suggesting that sedentary behavior associated with technology use may lead to weight gain.

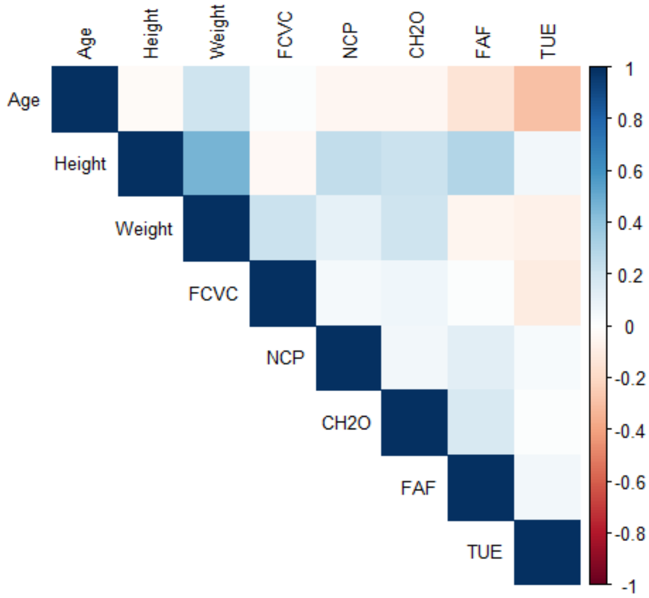

Figure 1. Correlation Matrix of Continuous Variables (Picture credit: Original)

In Figure 2, the bar plot demonstrates a correlation between family overweight history and levels of obesity. People with a family history of overweight are more likely to be classified as at high risk of obesity. This finding reinforces the notion of a genetic and familial predisposition to obesity, thereby underscoring the importance of considering family history in obesity risk assessments. It also highlights the potential benefit of early interventions in families with a history of obesity to mitigate future health risks.

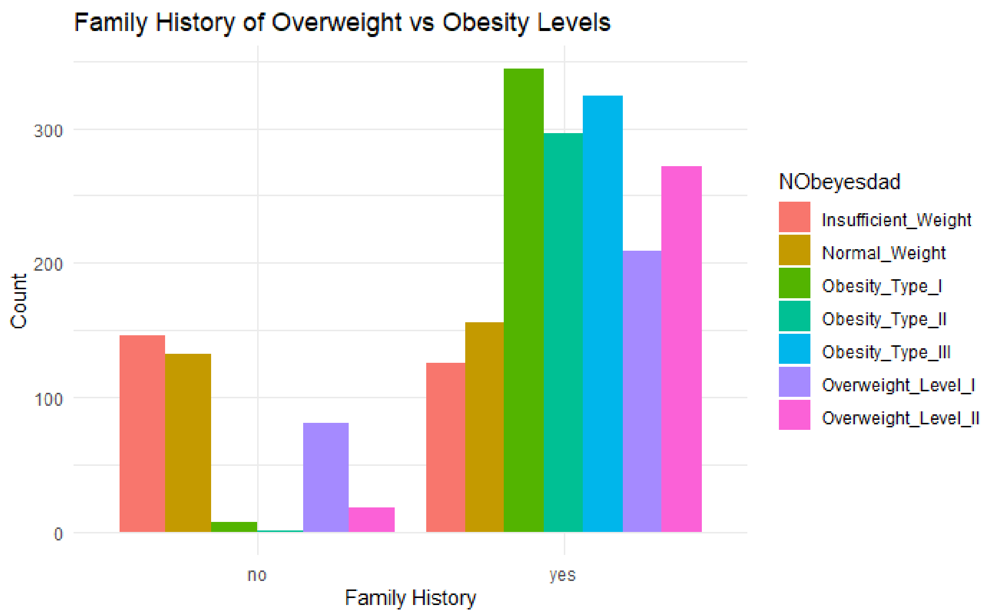

Figure 2. Family Overweight's History vs. Levels of Obesity (Picture credit: Original)

Figure 3 depicts the correlation between the frequency of physical activity and levels of obesity, as illustrated by a box plot. People who are "normal weight" engage in higher-intensity physical activity than people who are more obese. A notable decline in physical activity frequency is observed as obesity levels increase, indicating a correlation between lower physical activity and higher obesity levels. This finding corroborates the well-established correlation between a sedentary lifestyle and obesity, thereby underscoring the necessity of promoting regular physical activity as a pivotal element of obesity prevention and management initiatives.

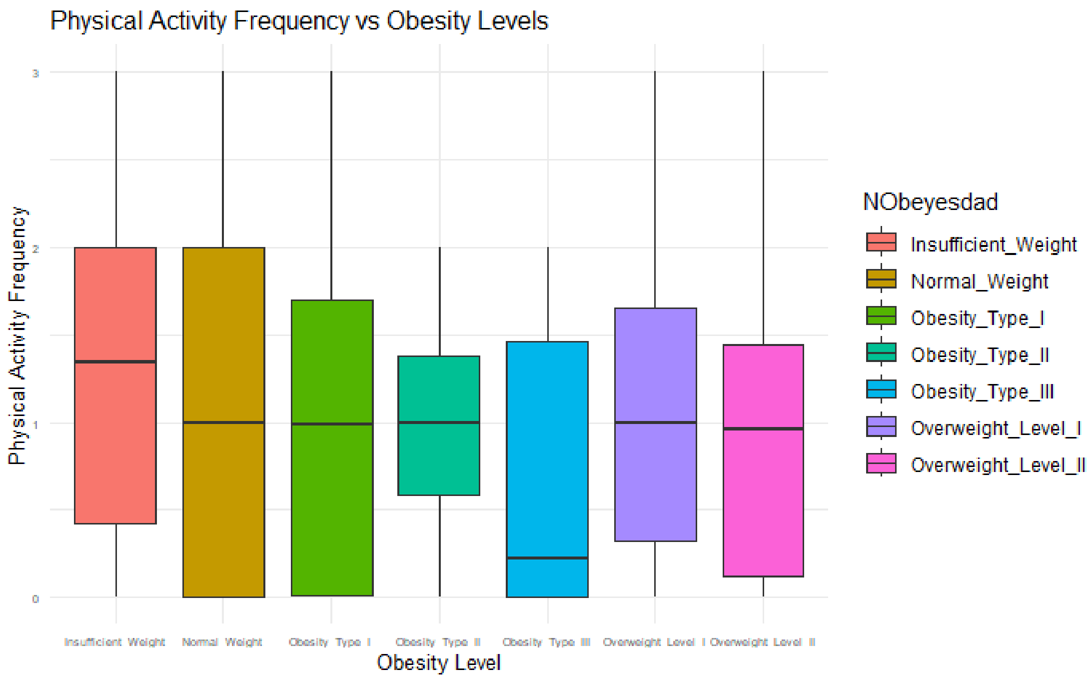

Figure 3. Physical Activity Frequency vs. Obesity Levels (Picture credit: Original)

Figure 4 illustrates the prevalence of vegetable consumption across distinct categories of obesity. Individuals classified as having "normal weight" tend to consume vegetables with greater frequency, whereas those in higher obesity categories generally consume less frequently. This pattern reinforces the importance of a diet rich in vegetables in maintaining a healthy weight. Encouraging the consumption of vegetables as part of a balanced diet could be an effective strategy for obesity prevention and management.

Figure 4. Vegetable Consumption Frequency vs. Obesity Levels (Picture credit: Original)

The ROC curves for each obesity level depicted in Figure 5 outline the performance of the logistic regression model in distinguishing between different obesity categories. These curves are mainly clustered near the upper left corner of the figure, indicating a high AUC value. It indicates that the model has a strong ability to distinguish between different obesity levels. This shows that there is a good balance between sensitivity and specificity, effectively predicting the various obesity categories in the dataset.

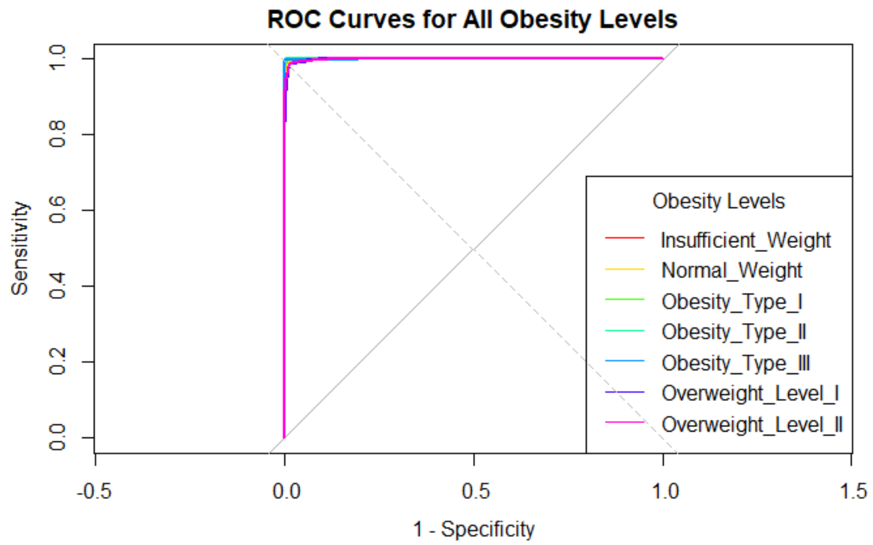

Figure 5. ROC Curve for All Obesity Levels (Picture credit: Original)

The result of this study contributes to a comprehension of the factors influencing obesity levels among the Latin American population, with a particular focus on Mexico, Peru, and Colombia. The logistic regression model exhibited a robust capacity to differentiate between obesity categories, particularly in the extremes of underweight and type III obesity. This suggests that the predictor variables incorporated into the model, including caloric intake, physical activity, and family history of overweight, are reliable indicators of these extremes. Nevertheless, difficulties were encountered when attempting to differentiate intermediate categories, such as level I in overweight and level II in overweight." The difficulty in classification may stem from the fact that individuals in these categories typically share similar lifestyle habits, such as caloric intake and physical activity levels. Overlapping characteristics make it difficult for the model to clearly distinguish between these groups, a challenge noted in similar obesity classification studies. The results of this study are consistent with previous research findings, emphasizing the impact of socioeconomic and behavioral factors on the development of obesity. The study indicates that time constraints, the accessibility of fast food, and the scarcity of opportunities for physical activity are significant contributors to obesity, particularly among lower-income populations. Furthermore, the Garcia team's article underscores the significance of family dynamics, particularly the impact of parental influence on their children's dietary habits [6]. The existing literature indicates that it is essential to engage the entire family in obesity prevention programs to achieve long-term success.

4. Conclusion

This study examined factors influencing individual obesity levels in Mexico, Peru, and Colombia, emphasizing the role of socioeconomic, behavioral, and genetic factors. EDA and LRM were both applied to identify key predictors of obesity. The results showed that variables such as physical activity, vegetable intake, and family obesity history had a significant impact on obesity levels, with clear differences between the extremes of “underweight” and “obesity type III”. However, the challenge of distinguishing closely related obesity levels such as “type I” and “type II” remained, suggesting that the addition or increase of more specific variables could improve model performance. ROC curve analysis further confirmed the strong predictive ability of the model. These results contribute to the understanding of the obesity problem in Latin America and highlight the need for comprehensive public health strategies that address not only dietary and physical activity habits but also the socioeconomic factors that contribute to obesity. The results of this study can inform future interventions and policies to try to reduce obesity rates and promote better lifestyles in the region.

References

[1] Anekwe CV, Jarrell AR, Townsend MJ, et al. Socioeconomics of Obesity. Curr Obes, 2020, 9 (3): 272 - 279.

[2] Gyu-Young L, Youn-Joo U. Factors affecting obesity in urban and rural adolescents: demographic, socioeconomic characteristics, health behavior and health education. International Journal of Environmental Research and Public Health, 2021, 18 (5): 2405.

[3] Fan X, Yun Z, Lingling Z, et al. Abdominal obesity: an independent influencing factor of visuospatial and executive/language ability and the serum levels of aβ40/aβ42/tau protein. Disease Markers, 2022: 3622149.

[4] Hossain S, Ankit G, Jonathan P, et al. Gpr75 -deficient mice are protected from high-fat diet–induced obesity. Obesity, 2023, 31 (4): 1024 – 37.

[5] Obesity Levels Data Set. Data in Brief 2019. https://doi.org/10.1016/j.dib.2019.104344.

[6] Garcia ML, Natalia G, Erika B, et al. Engaging intergenerational Hispanics/Latinos to examine factors influencing childhood obesity using the PRECEDE–PROCEED model. Maternal and Child Health Journal, 2019, 23 (6): 802 - 810.