# Cracking Obesity: Understanding the Interplay of Meal Patterns, Caloric Intake, and Snacking Habits*

Amos Okutse[1] Monica Colon-Vargas[1]
[1]School of Public Health, Brown University, Providence, RI.

**Abstract**

**Introduction:** In Latin America, obesity rates have reached alarming levels, posing serious health risks and placing a substantial burden on healthcare systems.

**Objective:** To decipher obesity dynamics focusing particularly on the triad interaction of main meals, high-calorie foods, and between-meal consumption rates.

**Methods:** We use cross-sectional data on $n = 498$ individuals aged between 14 and 61 years in three countries (Mexico, Peru, and Colombia) from the UC Irvine Machine Learning Repository. We categorize our outcome into normal weight, obese overweight, and underweight. Proportional odds logistic regression model fitted to covariates selected by both lasso penalization and significance based on preliminary analyses is used to decipher the complex interplay of factors in obesity. We use up-sampling to handle class imbalance and perform diagnostics to evaluate performance of the model.

**Results:** The significant interaction suggests that the impact of eating three complete meals daily on transitioning to a higher weight category is influenced by both consuming high-calorie foods and occasional between-meal eating. When both of these behaviors are present, the odds of moving to a higher weight category with each additional daily meal increase significantly, approximately 8.5 times higher compared to when these factors are absent.

**Conclusion:** We address the global health issue of obesity, with a specific focus on the diverse populations of Mexico, Peru, and Colombia. By examining individuals' eating habits and physical condition, tcontribute to the understanding of factors contributing to obesity prevalence in these regions and understand the underlying mechanisms driving the escalating rates of obesity.

***Keywords:*** *Obesity, Pandemic, Health risk, Lower socioeconomic groups.*

# Introduction

The World Health Organization (WHO) defines obesity as a chronic and intricate disease characterized by an accumulation of excess fat that poses significant health risks. This condition is not confined to isolated cases but has become alarmingly prevalent across numerous countries worldwide, warranting recognition as a global pandemic (Blüher 2019). Among its many complications, obesity

*Replication files are available on the author's Github account (https://github.com/MCV20/GLM-project.git).

**Current version**: April 28, 2024

significantly elevates the risk of various health issues, including diabetes, fatty liver disease, hypertension, cardiovascular events like heart attacks and strokes, cognitive decline, joint ailments such as osteoarthritis, disrupted sleep patterns due to conditions like obstructive sleep apnea, and an increased susceptibility to certain types of cancer (Blüher 2019).

Literature suggests that the likelihood of obesity is influenced by a range of factors beyond individual characteristics, including demographic attributes, community infrastructure, socioeconomic conditions, and specific environmental factors within communities (Gülü et al. 2022). In certain countries, particularly among lower socioeconomic groups, obesity rates have surged dramatically due to urbanization, shifts in diet and food availability, and reduced physical activity. This rise in obesity is linked to a significant increase in mortality from chronic diseases like type 2 diabetes, heart disease, and certain cancers, potentially shortening life expectancy by up to 20 years. Given the preventable nature of obesity and its associated health risks, early detection is crucial to mitigate the development of serious conditions such as cardiovascular issues, diabetes, and asthma. Obesity's complex origins involve various factors including socioeconomic status, occupation, and lifestyle habits like smoking and physical activity levels. Physical activity and eating habits are key in preventing obesity, as it primarily stems from an imbalance between calories consumed and expended. Weight loss typically involves reducing calorie intake, increasing energy expenditure, or both. When individuals consume more energy than needed, the excess is stored as fat, leading to obesity. Therefore, maintaining a healthy weight relies on a balanced diet and regular physical activity (Yagin et al. 2023).

This work is centered in identifying determinants associated with obesity, with particular emphasis on lifestyle behaviors. In Latin America, obesity rates have reached alarming levels, posing serious health risks and placing a substantial burden on healthcare systems (Popkin and Reardon 2018). Therefore this work will address the global health issue of obesity, with a specific focus on the diverse populations of Mexico, Peru, and Colombia. By examining individuals' eating habits and physical condition, this research aims to understand the factors contributing to obesity prevalence in these regions and understand the underlying mechanisms driving the escalating rates of obesity.

# Methods

## Data

The dataset, accessible online at the UC Irvine Machine Learning Repository, includes information pertinent to estimating obesity levels among $(n = 498)$ individuals aged 14 to 61 hailing from Mexico, Peru, and Colombia. Collected via a web-based survey administered to anonymous respondents, the dataset comprises 17 attributes and was available online for a duration of 30 days. Data collection involved posing questions as delineated in Table 1, which also details the variables under study and the methodology employed in the data collection process.

### Data Pre-processing

Given the relatively modest sample size of our dataset, we anticipate potential challenges in conducting the analyses. Our outcome variables comprise seven distinct levels, encompassing three tiers of obesity (Type I, II, III), two tiers of overweight (I, II), as well as normal weight and underweight categories. However, certain categories contain a limited number of individuals, necessitating the consoli-

Table 1: Variable Description

| Variable | Description | Values |
|---|---|---|
| `Nobeyesdad` | Obesity Level | Insufficient/Normal ObesityI/ObesityII ObesityIII/OverweightI/OverweightII |
| `Gender` | What is your gender? | Female/Male |
| `Age` | What is your age? | Numeric Value |
| `Height` | What is your height? | Numeric Value (m) |
| `Weight` | What is your weight? | Numeric Value (kg) |
| `family_history_with_overweight` | Has a family member suffered or suffers from overweight? | Yes/No |
| `FAVC` | Do you eat high caloric food frequently? | Yes/No |
| `FCVC` | Do you usually eat vegetables in your meals? | Never/Sometimes/Always |
| `NCP` | How many main meals do you have daily? | Numeric Value |
| `CAEC` | Do you eat any food between meals? | No/Sometimes/Frequently Always |
| `Smoker` | Do you smoke | Yes/No |
| `CH2O` | How much water do you drink daily | Numeric Value |
| `SCC` | Do you monitor the calories you eat daily | Yes/No |
| `FAF` | How often do you do physical activity | Numeric Value |
| `TUE` | How much time do you use technological devices such as cell phone, videogames, television, computer and others? | Numeric Value |
| `CALC` | How often do you drink alcohol | Never/Sometimes/ Frequently/Always |
| `MTRANS` | Which transportation do you usually use | Automobile/Motorbike Bike/Public/Walking |

dation of some for better statistical reliability. Accordingly, we introduce a novel class termed "Obese," which consolidates the three levels of obesity, alongside a separate class denoted as "Overweight," housing the two levels thereof, while retaining the remaining categories unchanged. Furthermore, certain categories within the covariates exhibited minimal representation, with some even lacking individuals in certain categories. Consequently, we undertook additional aggregation of these classes. For instance, within the alcohol consumption variable, we combined the "always" and "frequently" classes into a singular category. Similarly, in the transportation variable, we merged "motorbike" and "automobile" into a consolidated class termed "motor vehicle," while amalgamating "walking" and "bike" into a unified category.

## Statistical Modeling

Since the outcome variable exhibits an ordinal nature, where the categories can be ordered as underweight < normal < overweight < obese, we opted for a modeling approach suited to such data characteristics. Specifically, we employed the generalized linear model framework, fitting a proportional odds model. This modeling technique allows us to account for the ordinal nature of the outcome variable and the inherent ordering of its categories. By utilizing the proportional odds model, we can assess the relationship between the predictors and the ordinal outcome variable while accommodating the cumulative nature of the categories.

### Variable Selection

In light of the considerable number of variables present within the dataset, an approach to variable selection opted. To this end, we employed two distinct methodologies for variable selection. Primarily, variables were selected based on their significance as determined by p-values derived from the proportional odds model. Additionally, we conducted variable selection utilizing a proportional odds model with lasso penalization, a technique that inherently incorporates variable selection by penalizing less influential predictors. Subsequently, interactions among selected variables were explored, and the resultant models were meticulously compared to discern the most parsimonious and informative model structure.

### Class Imbalance

Given the class imbalance within the outcome variable, we address this issue by employing an upsampling technique. The goal is to decrease the disparity by augmenting the instances of minority classes to achieve parity with the majority class, thereby fostering a more balanced representation across all categories. Subsequently, within this augmented dataset, we applied the same two variable selection techniques previously described. This approach ensures that the variable selection procedures are conducted on a dataset that reflects a more equitable distribution of observations across the various outcome categories, thereby mitigating potential biases and bolstering the reliability of the ensuing models. We additionally proceed to fit an additional model using the augmented data set,incorporating solely those covariates that demonstrated statistical significance within the original data set.

## Model Diagnostics

**Likelihood Ratio Tests (LRT) and Information Criterion.** The majority of models analyzed in this study were nested, with assessments primarily relying on information criteria such as the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Notably, the BIC was preferred over the AIC due to its heightened penalty with larger sample sizes, leading to more stringent significance thresholds (Vrieze 2012). Model comparisons were conducted using the likelihood ratio test (LRT), aimed at examining competing models by testing the null hypothesis that the simpler and more complex models are equally effective. Specifically, this test assesses whether the additional parameters in the larger model significantly improve fit, implying that their effect sizes are statistically indistinguishable from zero.

**Proportional Odds Assumption.** We tested the parallel regression assumption of the fitted proportional odds model using the Brant test using the `brant` package and function of the same name in R. The Brant test of the parallel regression assumption tests the null hypothesis that the proportional odds assumption holds for all covariates combined and when considered independently.

**Goodness of fit test:** We used the chi-squared test to assess whether a statistically significant difference between the observed and predicted frequencies for each outcome class. Our null hypothesis using the chi-squared test for goodness of fit test was that there would be a significant difference between the observed and predicted outcome class frequencies, implying that the model fits the data poorly.

**Leverage, Influential Observations and Multicollinearity.** Leverage and influential observations in the data set were examined by deleting the $i^{th}$ observation and fitting the model without this observation. A change in the parameter estimates (in their absolute values) was then used as an indicator of the influence of observation $i$. We examined multicollinearity using the variance inflation factor (VIF), that is, a measure of how the variance of the estimated coefficients is increased due to multicollinearity.

**Variable Importance and Sensitivity Analysis.** We characterized variable importance by the magnitude of the coefficients. Larger coefficients (in absolute value) indicate more important variables. Sensitivity analysis involved assessing the robustness of the model estimate results to changes in modeling assumptions or data specifications. Here, we changed the fitting method to `probit` and to examine whether the rank of the important variables would change substantially. This would provide an idea of the stability of the variables.

# Results

## Descriptive Statistics

Table 2 presents an overview of participant characteristics stratified by each outcome category. Notably, the distribution across categories reveals a significant class imbalance, with only 34 participants categorized as underweight compared to 64 classified as obese. Conversely, the overweight category encompasses 116 participants, with the highest representation observed within the normal weight classification.

Table 2: Summary Statistic Stratified by Outcome Category

| Characteristic | **Overall**, N = 498 | **Underweight**, N = 34 | **Normal.Weight**, N = 287 | **Overweight**, N = 116 | **Obese**, N = 61 |
|---|---|---|---|---|---|
| **Gender** | | | | | |
| Female | 227.00 (45.58%) | 17.00 (50.00%) | 141.00 (49.13%) | 46.00 (39.66%) | 23.00 (37.70%) |
| Male | 271.00 (54.42%) | 17.00 (50.00%) | 146.00 (50.87%) | 70.00 (60.34%) | 38.00 (62.30%) |
| **Age** | 23.15 (6.72) | 20.15 (4.03) | 21.74 (5.10) | 25.77 (8.39) | 26.46 (8.22) |
| **Height** | 1.69 (0.10) | 1.70 (0.11) | 1.68 (0.09) | 1.69 (0.10) | 1.71 (0.10) |
| **Family history overweight** | 300.00 (60.24%) | 16.00 (47.06%) | 155.00 (54.01%) | 76.00 (65.52%) | 53.00 (86.89%) |
| **High-calorie food consumption** | 348.00 (69.88%) | 21.00 (61.76%) | 208.00 (72.47%) | 75.00 (64.66%) | 44.00 (72.13%) |
| **Number of meals veggie consumption** | | | | | |
| 1 | 32.00 (6.43%) | 3.00 (8.82%) | 18.00 (6.27%) | 7.00 (6.03%) | 4.00 (6.56%) |
| 2 | 272.00 (54.62%) | 12.00 (35.29%) | 155.00 (54.01%) | 69.00 (59.48%) | 36.00 (59.02%) |
| 3 | 194.00 (38.96%) | 19.00 (55.88%) | 114.00 (39.72%) | 40.00 (34.48%) | 21.00 (34.43%) |
| **Daily main meals** | | | | | |
| 1 | 108.00 (21.69%) | 5.00 (14.71%) | 52.00 (18.12%) | 35.00 (30.17%) | 16.00 (26.23%) |
| 3 | 344.00 (69.08%) | 21.00 (61.76%) | 206.00 (71.78%) | 73.00 (62.93%) | 44.00 (72.13%) |
| 4 | 46.00 (9.24%) | 8.00 (23.53%) | 29.00 (10.10%) | 8.00 (6.90%) | 1.00 (1.64%) |
| **Eating between meals** | | | | | |
| Frequently | 189.00 (37.95%) | 18.00 (52.94%) | 118.00 (41.11%) | 37.00 (31.90%) | 16.00 (26.23%) |
| no | 20.00 (4.02%) | 3.00 (8.82%) | 10.00 (3.48%) | 5.00 (4.31%) | 2.00 (3.28%) |
| Sometimes | 289.00 (58.03%) | 13.00 (38.24%) | 159.00 (55.40%) | 74.00 (63.79%) | 43.00 (70.49%) |
| **Smoke** | 32.00 (6.43%) | 1.00 (2.94%) | 13.00 (4.53%) | 8.00 (6.90%) | 10.00 (16.39%) |
| **Water Intake (L)** | | | | | |
| 1 | 135.00 (27.11%) | 10.00 (29.41%) | 83.00 (28.92%) | 25.00 (21.55%) | 17.00 (27.87%) |
| 2 | 266.00 (53.41%) | 16.00 (47.06%) | 164.00 (57.14%) | 61.00 (52.59%) | 25.00 (40.98%) |
| 3 | 97.00 (19.48%) | 8.00 (23.53%) | 40.00 (13.94%) | 30.00 (25.86%) | 19.00 (31.15%) |
| **Calorie intake monitoring** | 55.00 (11.04%) | 6.00 (17.65%) | 30.00 (10.45%) | 16.00 (13.79%) | 3.00 (4.92%) |
| **Frequency of days of physical activity (wk)** | | | | | |
| 0 | 162.00 (32.53%) | 10.00 (29.41%) | 80.00 (27.87%) | 44.00 (37.93%) | 28.00 (45.90%) |
| 1 | 158.00 (31.73%) | 6.00 (17.65%) | 97.00 (33.80%) | 40.00 (34.48%) | 15.00 (24.59%) |
| 2 | 113.00 (22.69%) | 14.00 (41.18%) | 69.00 (24.04%) | 18.00 (15.52%) | 12.00 (19.67%) |
| 3 | 65.00 (13.05%) | 4.00 (11.76%) | 41.00 (14.29%) | 14.00 (12.07%) | 6.00 (9.84%) |
| **Technology use time** | | | | | |
| 0 | 243.00 (48.80%) | 13.00 (38.24%) | 129.00 (44.95%) | 68.00 (58.62%) | 33.00 (54.10%) |
| 1 | 181.00 (36.35%) | 13.00 (38.24%) | 122.00 (42.51%) | 30.00 (25.86%) | 16.00 (26.23%) |
| 2 | 74.00 (14.86%) | 8.00 (23.53%) | 36.00 (12.54%) | 18.00 (15.52%) | 12.00 (19.67%) |
| **Frequency of alchohol intake** | | | | | |
| Frequently | 46.00 (9.24%) | 1.00 (2.94%) | 19.00 (6.62%) | 17.00 (14.66%) | 9.00 (14.75%) |
| no | 179.00 (35.94%) | 14.00 (41.18%) | 107.00 (37.28%) | 36.00 (31.03%) | 22.00 (36.07%) |
| Sometimes | 273.00 (54.82%) | 19.00 (55.88%) | 161.00 (56.10%) | 63.00 (54.31%) | 30.00 (49.18%) |
| **Transportation** | | | | | |
| Motor_Vehicle | 110.00 (22.09%) | 3.00 (8.82%) | 51.00 (17.77%) | 34.00 (29.31%) | 22.00 (36.07%) |
| Public_Transportation | 326.00 (65.46%) | 25.00 (73.53%) | 200.00 (69.69%) | 66.00 (56.90%) | 35.00 (57.38%) |
| Walking/Bike | 62.00 (12.45%) | 6.00 (17.65%) | 36.00 (12.54%) | 16.00 (13.79%) | 4.00 (6.56%) |

[1] n (%); Mean (SD)

Within the subset of 64 obese participants, 53 (86.80%) individuals exhibit familial overweight history, whereas among the underweight subset, 16 (47.06%) individuals demonstrate a similar familial predisposition. Regarding dietary habits, individuals categorized as normal weight exhibit the highest prevalence of high-caloric food consumption, followed sequentially by obese, overweight, and underweight counterparts. Additionally, observing the frequency of vegetable consumption, we can see 55.88% of underweight participants consuming vegetables in all three main meals, while 59% of obese and overweight individuals consume vegetables in two main meals. Meal frequency reveals that most participants consume three daily main meals, while a minority consumes only one daily meal. Moreover, between-meal eating habits are prevalent across all weight categories, with few individuals abstaining from between-meal consumption. Remarkably, a significant proportion (52.95%) of underweight participants report frequent between-meal snacking. Hydration patterns among participants indicate a consistent consumption of approximately 2-3 liters of water daily. Public transportation emerges as the preferred mode of transportation among participants. Furthermore, physical activity engagement varies across weight categories, with 45.9% of obese individuals reporting no weekly physical activity.

## Model Comparisons

A proportional odds model was employed to analyze the original dataset comprising $(n = 498)$ observations. Variable selection techniques, specifically LASSO regularization and p-value-based selection, yielded identical sets of significant predictors. These predictors included age, height, daily number of complete meals, weekly physical activity, daily water intake (L), and technology usage duration. To comprehensively explore potential interactions among these predictors, two-way and three-way interaction terms were examined. The interaction between age and height, the daily frequency of meals and physical activity, as well as physical activity and water intake, were found to be statistically non-significant. However, a three-way interaction involving the daily frequency of meals, physical activity, and water intake exhibited significance across most coefficient estimates. Despite this, a series of likelihood ratio tests were conducted, comparing models with varying complexity to a simpler baseline model. Likelihood ratio tests consistently favored the simpler model over those incorporating the interaction terms. Consequently, despite the significant findings regarding the three-way interaction, the simpler model without interactions was deemed more favorable based on the statistical evidence obtained from the likelihood ratio tests. Table 3 displays the AIC and BIC values for the various models. It is evident that the reduced model, which excludes interactions, exhibits the lowest AIC and BIC scores.

Table 3: AIC and BIC

| | AIC | BIC |
|---|---|---|
| mod.polr.reduced | 1029.151 | 1088.099 |
| mod1 | 1031.062 | 1094.221 |
| mod2 | 1038.486 | 1122.698 |
| mod3 | 1037.142 | 1121.354 |
| mod4 | 1055.112 | 1223.536 |

One important issue observed in these models pertains to class imbalance, where a disproportionate focus on these models is placed on accurately predicting instances within the normal weight class, while neglecting predictions for other weight categories. This imbalance is evidenced by the extreme case where the obese class fails to yield any predicted instances. Consequently, we address this imbalance by generating a new synthetic dataset wherein the underrepresented classes were upsampled. This augmented dataset now encompasses a total of $n = 1148$ observations, evenly distributed with 287 observations allocated to each outcome category.

In the augmented dataset, both variable selection methods are employed. The LASSO penalty method exhibits limited variable selection, as it only omits the gender variable. Conversely, selection based on p-values results in a more concise set of significant variables, including family history of overweight, high-calorie food consumption, daily number of complete meals, eating between meals, physical activity, and technology usage duration.

Two-way and three-way interactions were considered, including several combinations. Specifically, these interactions involved high-calorie food consumption with daily main meals, the number of daily main meals combined with eating between meals, physical activity in conjunction with the number of daily main meals, daily main meals coupled with physical activity, physical activity paired with eating between meals, and lastly, the three-way interaction among the number of daily meals, physical activity, and eating between meals. It is noteworthy that most interaction terms demonstrated a high level of significance within the models. We conducted comparisons between the simple model and the model incorporating interactions, affirming the significance of these interactions through likelihood ratio tests. Additionally, when comparing models with interactions, the Likelihood ratio test indicated a preference for the model featuring the three-way interaction by concluding that its fit significantly differs from that of the others. Table 4 presents the AIC and BIC, revealing that the model including three interactions (`mod.up.int5`) exhibits the lowest AIC and BIC scores.

Table 4: AIC and BIC

| | AIC | BIC |
|---|---|---|
| mod.up.reduced | 2918.998 | 2989.639 |
| mod.up.int1 | 2878.510 | 2959.242 |
| mod.up.int2 | 2859.299 | 2945.077 |
| mod.up.int3 | 2887.828 | 2988.744 |
| mod.up.int4 | 2893.649 | 2994.565 |
| mod.up.int5 | 2801.186 | 2922.284 |

## Unveiling the complex interplay between factors determining obesity

In Table 5, we present significant coefficients of the model with three way interactions. The statistically significant three-way interaction between `NCP3`, `FAVCyes`, and `CAECSometimes` suggests that the effect of eating three complete main meals daily on the odds of being in a higher weight category compared to a lower one is modified by the presence of both eating high calorie food consumption and sometimes eating between meals. When both eating high caloric food and sometimes eating between meals are present, the odds of moving to a higher weight category for each unit increase in number of daily meals are approximately 8.5 (e^2.142) times higher compared to when these variables are absent.

## Diagnostics and Performance Evaluation

We present results of our diagnostics and performance evaluation for the selected model. This model included the three-way interaction between number of main meals per day, frequent consumption of high calorie foods and the rate consumption of food between meals.

Our results based on the Brant test of the parallel regression assumption did not indicate substantial deviations of our selected model from the parallel regression assumption. We found no significant evidence of the combined effect of all covariates deviating from the null hypothesis ($\chi^2_{(22)} = 1022.54, p < 0.001$).

Table 5: Regression coefficients based on the proportional odds logistic model

| | estimate | std.errors | statistic | pvalues |
|---|---|---|---|---|
| fam.histyes | 1.323 | 0.353 | 3.751 | 0.0002 |
| FAF1 | -0.406 | 0.375 | -1.082 | 0.2794 |
| FAF2 | -0.439 | 0.409 | -1.073 | 0.2832 |
| FAF3 | -0.568 | 0.438 | -1.297 | 0.1947 |
| NCP3 | 2.671 | 0.831 | 3.214 | 0.0013 |
| NCP4 | 2.995 | 0.841 | 3.559 | 0.0004 |
| FAVCyes | 2.840 | 0.828 | 3.429 | 0.0006 |
| CAECno | 1.913 | 1.283 | 1.492 | 0.1357 |
| CAECSometimes | 3.585 | 0.822 | 4.361 | 0.0000 |
| TUE1 | -0.607 | 0.364 | -1.668 | 0.0953 |
| TUE2 | -0.317 | 0.409 | -0.775 | 0.4383 |
| NCP3:FAVCyes | -2.407 | 0.869 | -2.770 | 0.0056 |
| NCP4:FAVCyes | -4.128 | 0.904 | -4.568 | 0.0000 |
| NCP3:CAECno | -4.081 | 1.321 | -3.089 | 0.0020 |
| NCP3:CAECSometimes | -2.796 | 0.863 | -3.241 | 0.0012 |
| NCP4:CAECSometimes | -2.920 | 1.299 | -2.249 | 0.0245 |
| FAVCyes:CAECno | -0.113 | 1.310 | -0.086 | 0.9314 |
| FAVCyes:CAECSometimes | -2.305 | 0.865 | -2.664 | 0.0077 |
| NCP3:FAVCyes:CAECno | 3.941 | 1.368 | 2.881 | 0.0040 |
| NCP3:FAVCyes:CAECSometimes | 2.142 | 0.911 | 2.350 | 0.0188 |
| NCP4:FAVCyes:CAECSometimes | 0.101 | 1.337 | 0.076 | 0.9396 |
| Underweight\|Normal.Weight | 2.053 | 0.785 | 2.616 | 0.0089 |
| Normal.Weight\|Overweight | 3.540 | 0.788 | 4.492 | 0.0000 |
| Overweight\|Obese | 4.912 | 0.790 | 6.214 | 0.0000 |

Our chi-square goodness of fit test results provided evidence against the null hypothesis of poor fit alluding to the model performing relatively well. The predicted frequencies for each outcome appeared to be within what would be expected under a good fit ($\chi^2_{(3)} = 232.33, p < 0.001$).

Influential observation check did not reveal any substantial issues. The maximum change in any parameter estimate when each observation is left out was no more than 1.38 (on the log odds scale). Observations that cause a large change in the estimates might be considered influential with a common rule of thumb being that an observation might be influential if removing it changes the coefficient estimate by more than 20%. Additionally, there were no substantial issues with collinearity with all VIF Values being less than. VIF values up to 5 are usually interpreted as uncritical, values above 5 denote a considerable multicollinearity (Signorell 2024).

Figure 1 depicts the variable importance under the proportional odds logistic regression. The model shows the significant role of the interaction between number of main meals per day (three meals), frequent consumption of high calorie foods (yes) and the rate consumption of food between meals (no snacking). This is followed by the interaction between the number of main meals per day (three meals) and no snacking.

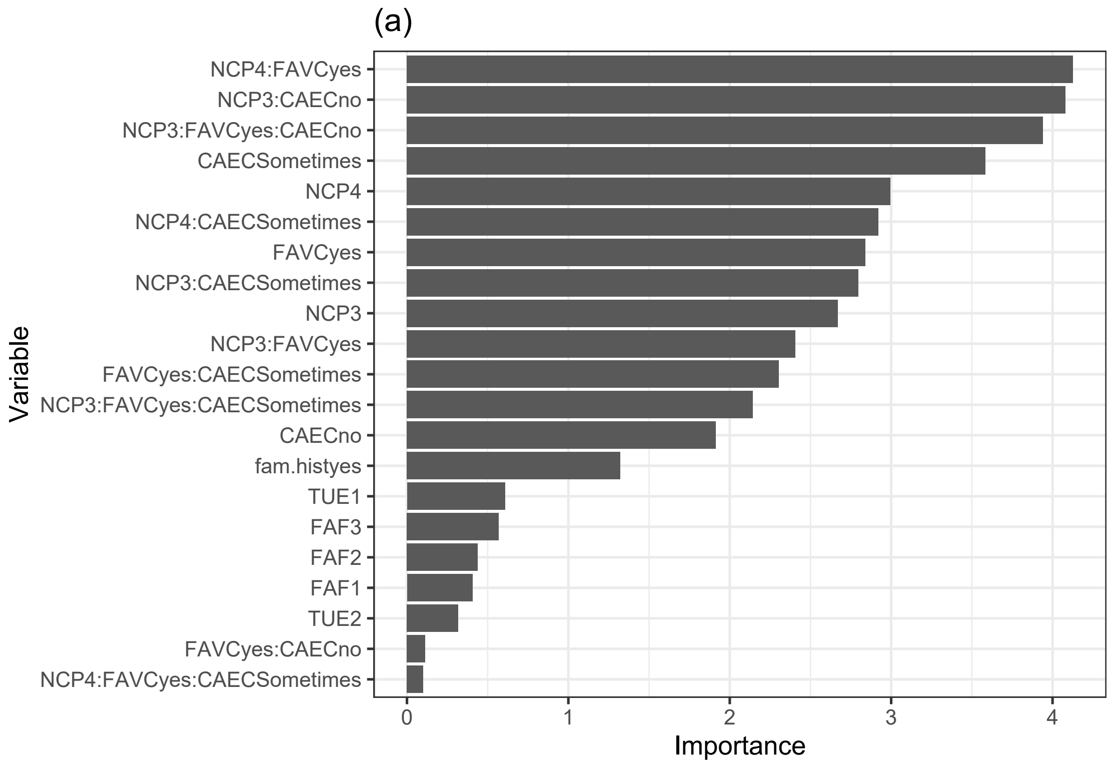


Figure 1: Variable importance under the proportional odds logistic regression

In Figure 2, we compare these covariates and those ranked by the probit regression method of fitting the model. We see that the first 7 coefficients rank similarly when the two methods are used and the results only slightly vary. Sensitivity analysis revealed (to some extent) the robustness of the model estimate results to changes in modeling assumptions providing an idea of the stability of the variables. The slight differences in how these models rank the coefficients could suggest that the relationship between that predictor and the response is not stable across different modeling approaches. Inconsistent rankings of coefficients between models in sensitivity analysis highlight the need for cautious interpretation and further investigation to ensure robust and reliable model results.

# Discussion

## Overview

The objective of this study is to ascertain the determinants that lead to obesity, with a particular emphasis on the interplay between socioeconomic indicators and lifestyle behaviors across the diverse demographics of Mexico, Peru, and Colombia. The alarming rates of obesity in Latin America pose a major health challenge and place a significant strain on healthcare systems. This investigation endeavors to gain insight into the factors that contribute to the high prevalence of obesity in these areas,

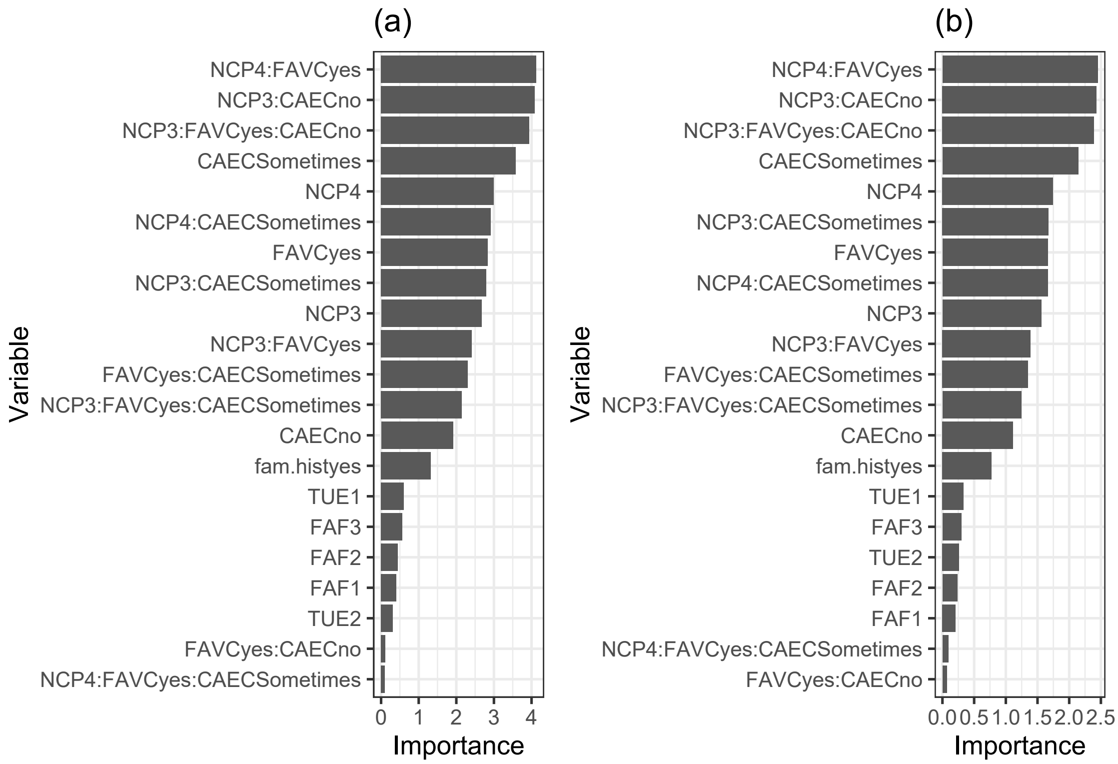


Figure 2: Variable importance under the proportional odds logistic regression

as well as the underlying mechanisms that propel the escalating rates of obesity. To achieve this, the study will delve into individuals' dietary patterns and physical well-being.

## Key Finding

The proportional logistic regression model yielded noteworthy results, uncovering several significant predictors of obesity. Specifically, the variables `NCP3`, `NCP4`, `FAVCyes`, `CAECno`, `CAECSometimes`, and their interactions were found to be considerably associated with obesity. These findings underscore the critical impact that these factors - which reflect different facets of a person's dietary habits and physical state - have on determining levels of obesity.

## Study strengths

An important advantage of this research is its emphasis on the varied populations of Mexico, Peru, and Colombia, which offers a broader perspective on the factors contributing to obesity in these regions. Moreover, the proportional logistic regression model employed in the study facilitated the investigation of the interrelationship between different socioeconomic factors and lifestyle choices, yielding meaningful insights into the intricate dynamics underlying obesity prevalence.

The finding of a significant three-way interaction between the number of main meals per day, frequent consumption of high-calorie foods, and the rate of consumption of food between meals in relation to the obesity complex is highly substantive in the context of public health, particularly in countries like

Mexico, Peru, and Colombia, as well as globally.

Similar to many countries worldwide, Mexico, Peru, and Colombia are facing an obesity epidemic. It is essential to comprehend the intricate factors that contribute to obesity to design public health interventions that are effective in addressing this issue. The interactions that have been identified reveal particular dietary behaviors that may either worsen or alleviate the risk of obesity. These insights can be invaluable in crafting targeted interventions.

Lasting, while the study focused on Mexico, Peru, and Colombia, the identified interactions likely have broader relevance beyond these countries. Similar patterns of dietary behaviors and their associations with obesity risk are observed globally. Therefore, understanding these interactions can inform public health efforts in diverse cultural and geographical settings.

## Limitations and Future Directions

Despite this study's strengths, some limitations must be considered. One is that self-reporting was used to collect the data, which could be subject to reporting bias. To validate the findings, future studies should include objective measures of physical activity and dietary intake.

Another limitation is that the study was cross-sectional, which means that causality cannot be inferred. To confirm the causal relationships suggested by the findings, longitudinal studies are necessary.

Lastly, while the study accounted for several important factors, other unmeasured confounding variables could still influence the results. Future research should consider additional factors, such as genetic predispositions, mental health status, and environmental influences.

## Conclusion

In conclusion, this study offers valuable insights into the factors that contribute to obesity in the diverse populations of Mexico, Peru, and Colombia. The findings underscore the significance of lifestyle behaviors in relation to obesity and emphasize the need for interventions that promote healthy eating habits and regular physical activity. Future research should address the limitations of the current study and continue to explore the intricate interplay between socioeconomic indicators and lifestyle behaviors in relation to obesity.

# References


Blüher, Matthias. 2019. "Obesity: Global Epidemiology and Pathogenesis." *Nature Reviews Endocrinology* 15 (5): 288–98.

Gülü, Mehmet, Hakan Yapici, Elena Mainer-Pardos, Ana Ruivo Alves, and Hadi Nobari. 2022. "Investigation of Obesity, Eating Behaviors and Physical Activity Levels Living in Rural and Urban Areas During the Covid-19 Pandemic Era: A Study of Turkish Adolescent." *BMC Pediatrics* 22 (1): 405.

Popkin, Barry M, and Thomas Reardon. 2018. "Obesity and the Food System Transformation in Latin America." *Obesity Reviews* 19 (8): 1028–64.

Signorell, Andri. 2024. *DescTools: Tools for Descriptive Statistics*. https://CRAN.R-project.org/package=DescTools.

Vrieze, Scott I. 2012. "Model Selection and Psychological Theory: A Discussion of the Differences Between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)." *Psychological Methods* 17 (2): 228.

Yagin, Fatma Hilal, Mehmet Gülü, Yasin Gormez, Arkaitz Castañeda-Babarro, Cemil Colak, Gianpiero Greco, Francesco Fischetti, and Stefania Cataldi. 2023. "Estimation of Obesity Levels with a Trained Neural Network Approach Optimized by the Bayesian Technique." *Applied Sciences* 13 (6): 3875.