Nicholas Mata
jnm705

Background

Obesity and overweight are significant health concerns marked by abnormal accumulation of body fat, increasing the risk of various chronic illnesses such as diabetes, heart disease, and certain cancers. Just one side of malnutrition, obesity has surpassed underweight in prevalence across most parts of the world, except South-East Asia. Obesity is influenced by a range of factors, including genetics, lifestyle choices, and behavioral patterns like diet and physical activity. Understanding these factors is essential for developing effective strategies to manage and prevent obesity while promoting overall nutritional well-being.

Motivation (and questions to answer/explore)

Health and longevity have their importance for individuals, whether they are part of the aging population, the active workforce, or just entering their adult lives. This analysis will give insight into commonalities between persons in different obesity levels, as well as enable them to make informed decisions about their lifestyle choices. One's demographics, lifestyle habits, and family history will be explored to find the relationships between them and obesity.

The analysis will explore the following:

Demographics

- How does obesity vary across different genders and age groups?
- Is there a correlation between age and BMI?

Lifestyle and Eating Habits

- Does the frequency of vegetable consumption (FCVC) correlate with lower obesity levels?
- Do people who frequently eat between meals (CAEC) show higher obesity levels?
- Is there a relationship between the number of daily meals (NCP) and obesity level?
- Is there a significant difference in BMI between smokers and non-smokers?
- How do water and alcohol intake individually and combined affect obesity level?

Family History

- Are individuals with a family history of being overweight more likely to be obese?
- How does family history influence BMI when combined with other factors like high-calorie food consumption or smoking?

Lastly, the most significant factors that influence obesity level (NObeyesdad) will be identified and used to create a predictive model for obesity level.

Data Description

The dataset includes data to estimate obesity levels in individuals from Mexico, Peru, and Colombia. 77% of the available data was generated synthetically and the remaining 23% was collected from users via web platforms. Eating and lifestyle habits, physical condition, and family history information was recorded totaling 17 attributes from 2111 records. Each record is labeled with the class variable NObeyesdad (obesity level) that classifies the data with values of Insufficient Weight, Normal Weight, Overweight I, Overweight II, Obesity Type I, Obesity Type II, and Obesity Type III. The data procured was relatively clean with no missing values throughout, and contains a mix of categorical and numerical variables.

Proposed Analysis

Going forward, body mass index (BMI) will be calculated and appended to each case. I believe this addition will aid in the analysis and prediction. The analysis will include logistic and linear regressions, statistical tests such as ANOVA and Chi-Square, and graphs including histograms and box plots. Applying these methods, and others, will give insight into the factors that influence obesity the most. Once these most influential variables are found, a model will be fit to predict the category of obesity for the individuals. Ordinal logistic regression would be the best for this because our response variable is levels of obesity. A multinomial logistic regression will also be explored but is expected to perform suboptimally.

References

Cleveland Clinic medical. “BMI (Body Mass Index): What It Is & How to Calculate.” Cleveland Clinic, 18 Feb. 2025, my.clevelandclinic.org/health/articles/9464-body-mass-index-bmi.

Estimation of Obesity Levels Based On Eating Habits and Physical Condition [Dataset]. (2019). UCI Machine Learning Repository. https://doi.org/10.24432/C5H31Z.

“Obesity.” World Health Organization, World Health Organization, 2025, www.who.int/health-topics/obesity#tab=tab_1.

Appendix

The raw data: [link]

Estimation of obesity levels in individuals from Mexico, Peru, and Colombia based on their eating habits and physical condition.

The data was from a csv containing 17 attributes and 2111 records.
The variables are described as follows:

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Gender	Feature	Categorical	Gender			no
Age	Feature	Continuous	Age			no
Height	Feature	Continuous				no
Weight	Feature	Continuous				no
family_history_with_overweight	Feature	Binary		Has a family member suffered or suffers from overweight?		no
FAVC	Feature	Binary		Do you eat high caloric food frequently?		no
FCVC	Feature	Integer		Do you usually eat vegetables in your meals?		no
NCP	Feature	Continuous		How many main meals do you have daily?		no

CAEC	Feature	Categorical	Do you eat any food between meals?	no
SMOKE	Feature	Binary	Do you smoke?	no
CH2O	Feature	Continuous	How much water do you drink daily?	no
SCC	Feature	Binary	Do you monitor the calories you eat daily?	no
FAF	Feature	Continuous	How often do you have physical activity?	no
TUE	Feature	Integer	How much time do you use technological devices such as cell phone, videogames, television, computer and others?	no
CALC	Feature	Categorical	How often do you drink alcohol?	no
MTRANS	Feature	Categorical	Which transportation do you usually use?	no
NObeyesdad	Target	Categorical	Obesity level	no