Obesity Unveiled: Insights from NHANES Data

Course ID: DS2500, Section 01

Instructor: Rushit Sanghrajka

Due Date: Dec. 3rd. 2024

Project Topic

- Exploring Factors Influencing Obesity Among U.S. Adults

Team Members

1. Yifan Wang, wang.yifan17@northeastern.edu
2. Guo Tang, tang.guo1@northeastern.edu
3. Yanzhen Chen, chen.yanzh@northeastern.edu
4. Siyuan Chen, chen.siyuan5@northeastern.edu

Project Background

Obesity and overweight are increasingly prevalent in today's society, especially in the United States, and their impact has caused widespread concern. According to statistics, about 40% of U.S. adults are obese and another 30% are overweight. Obesity not only poses a serious threat to individual health, such as increasing the risk of heart disease, diabetes and other chronic diseases, but also brings great pressure on the social health care system and economic development.

In order to effectively address this public health challenge, it is critical to gain a deeper understanding of obesity prevalence trends and the influencing factors behind them (e.g., age, gender, race, and socioeconomic status). The U.S. National Health and Nutrition Examination Survey (NHANES) provides detailed and authoritative data support, allowing us to analyze the complexity of the obesity problem from a data perspective.

Based on the NHANES data, this project examines the distribution of different weight categories (normal weight, overweight, and obese) among adults aged 20 years and older, with the aim of exploring the association between demographic characteristics and weight status. This analysis will not only help understand the key drivers of obesity, but also provide a scientific basis for policy makers to design precise health interventions.

Introduce to Our Data

- Data Source
    - We use data primarily from publicly available data from the U.S. government portal Data.gov, specifically the National Health and Nutrition Examination Survey's (NHANES) data.
    - Link:https://drive.google.com/file/d/1uj7qfUmwoX0gYYAQ1vp8jgFQnaTnTlel/view
- Shortcomings of data sources
    - Sampling bias:
        - The data are based on sample surveys, and there may be sample representation issues, such as underrepresentation of low-income groups, the homeless, or certain ethnic minorities.
    - Time span and trends:
        - Data covers multiple years but may not reflect recent trends.
    - Degree of refinement of demographic characteristics:
        - Indicators of socioeconomic status may be limited, although information such as age, gender, and ethnicity are included.
- Potential privacy or ethical issues
    - Interpretation and use of data:
        - Careful use of data is needed to avoid stigmatization of certain groups as a result of statistical findings
    - Policy Implications:

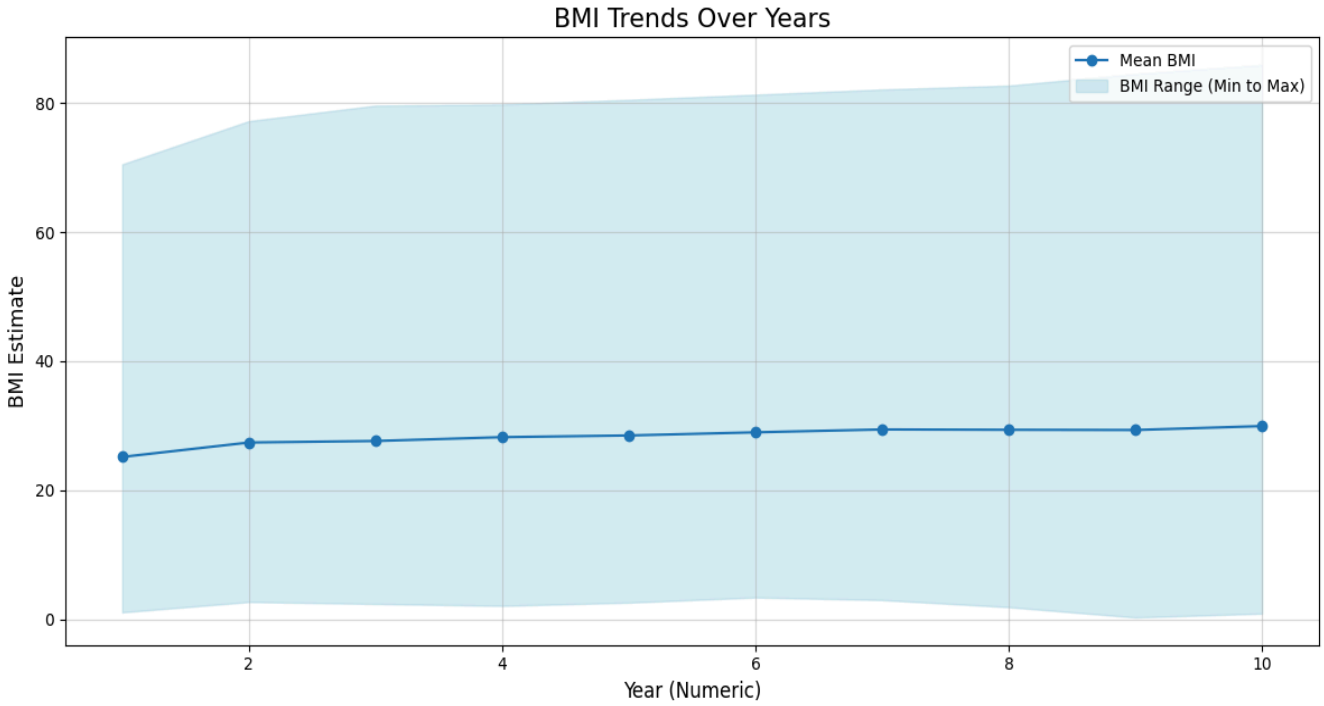
        - The results of the study may be used to formulate policies that, if not analyzed comprehensively enough, may overlook the health needs of certain groups.
- Advantages of information sources
    - Authoritative:
        - NHANES data is collected by the National Center for Health Statistics, the authoritative source for health and nutrition research.
    - Enrichment:
        - The data cover a wide range of demographic characteristics as well as detailed weight distribution information, making it suitable for in-depth analysis.
    - Comparability:
        - Data are grouped by year, supporting comparisons and trend analysis across years.

Data Science Approaches

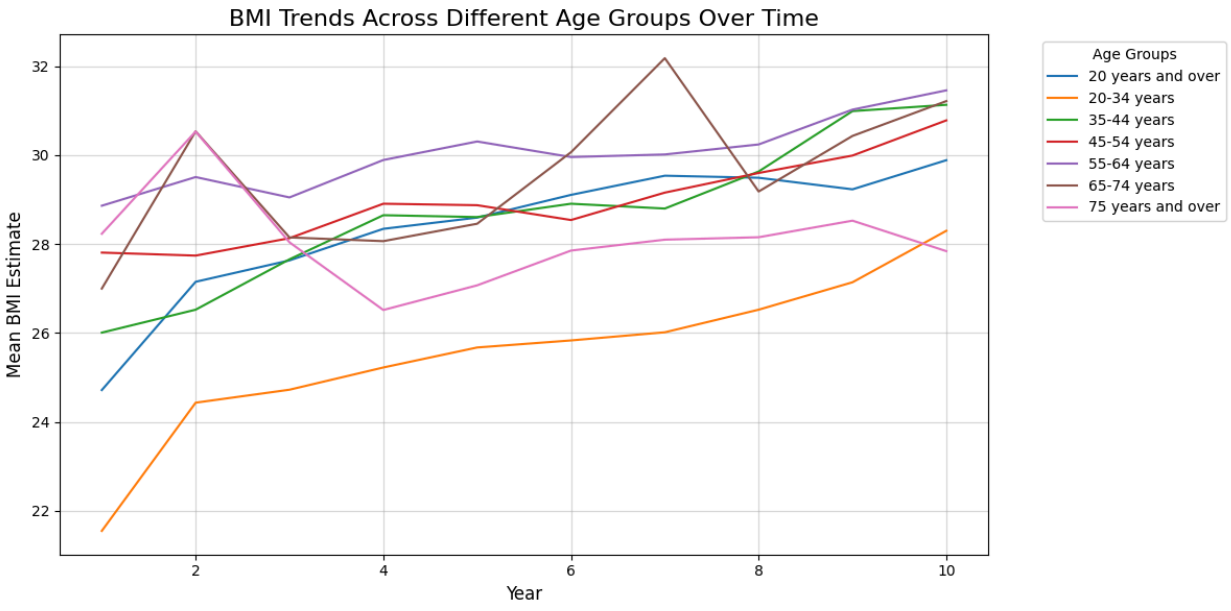
- We chose BMI (Body Mass Index) as our main observation, which is calculated by dividing the weight by the square of the height, and can be a good response to whether a person's weight is within the normal range.
- Descriptive statistics:
    - To analyze the underlying distribution of weight categories across demographic characteristics.
- Correlation analysis:
    - To assess the relationship between weight status and demographic characteristics.
- Logistic regression:
    - Predicts the likelihood of an individual being in an overweight or obese category.
- Multiple linear regression:
    - Assesses the effect of continuous variables (e.g., age, income) on weight indicators (e.g., BMI).
- Data visualization:
    - Show trends and relationships using histograms, bar graphs, scatter plots, etc.

Results and Conclusion

- **Trends of BMI (Weight(kg) / Height(m)^2) Categories Over Years:**
    - Based on what we can see in this graph, we can see that the average BMI of people is currently on an upward trend, meaning that people are facing a growing obesity problem, i.e., the ratio of higher weight to the square of height is becoming higher. At the same time, the highest values of BMI are on an increasing trend, meaning that a minority of the population is facing a very exaggerated obesity problem.
- **Relevant recommendations:**
    - There is a need to focus on obesity and overweight and to develop educational campaigns on healthy eating and exercise.
    - Medical support and weight loss programs are recommended for people with severe obesity.



- **BMI (Weight / Height^2) Trends Across Different Age Groups Over Time**
    - As you can see from this chart, the average BMI for all age groups is on an upward trend. The more noteworthy ones are:
        - Rapid rise in BMI among young people (20-34 years old).
        - Middle-aged adults (35-64 years) have a higher but relatively stable BMI.
        - BMI fluctuations in older age groups (65 years and older). Although the mean BMI was high initially, it gradually decreased thereafter.
- **Relevant recommendations:**
    - For young people: promote healthy eating habits and daily exercise to control excessive BMI growth.
    - For middle-aged adults: Increase awareness of weight management to prevent obesity-related chronic diseases (e.g., diabetes, hypertension).
    - For the senior population: focus on weight loss in the senior population to avoid malnutrition or bone health problems caused by low weight.



Future Work:

- Further data mining and analysis
    - Incorporate additional variables: integrate lifestyle data (e.g., diet, exercise), and mental health indicators (e.g., stress levels, sleep quality) into the analysis to explore their association with weight status.
- Interdisciplinary Research
    - Combining sociological studies: to explore socio-cultural influences on weight perception and obesity behaviors, e.g., differences in acceptance of obesity by race/culture.
- Public Health Applications
    - Targeted Population Interventions: Use the results of the analysis to develop personalized health intervention strategies for high-risk populations。
- Data Extension
    - External Data Integration: Integration of other publicly available data sources (e.g., CDC data from the U.S. Centers for Disease Control and Prevention, local health surveys) to improve analysis accuracy and data representativeness.

Reference Page

Centers for Disease Control and Prevention (CDC). (2019). *Health, United States, 2019 appendixes*. National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/hus/hus19-appendix-508.pdf

National Health and Nutrition Examination Survey (NHANES). (n.d.). *National health data*. Data.gov. Retrieved from https://www.data.gov