ALY6010 - Probability Theory and Introductory Statistics
Final Project

Factors Influencing Obesity in the United States: A Demographic and Lifestyle Analysis

Name: Hari Priya Ramamoorthy

NUID: 002324226

Date: October 31, 2024

Introduction

This study examines the influence of demographic and lifestyle factors on obesity rates in the United States—a pressing public health issue closely tied to chronic diseases, including heart disease, diabetes, and certain cancers. Understanding the demographic and lifestyle contributors to obesity is essential for policymakers to design effective public health interventions. To support this, the study leverages data from the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) to inform evidence-based policies that address obesity's root causes, helping federal and state agencies deploy resources more effectively and foster healthier communities.

The analysis proceeds in two main parts. First, an exploratory data analysis (EDA) of the BRFSS dataset, accessible via the U.S. Government's Open Data portal, identifies key statistical questions about obesity trends among the U.S. adult population. In the second part, T-tests, correlation, and regression analyses are used to evaluate these questions and determine the significance of various demographic and lifestyle factors in predicting obesity. Let's dive into analysis and insights.

Exploratory Data Analysis (EDA)

To understand the obesity trends, an exploratory data analysis was carried out with BRFSS dataset. Key variables analyzed include age, education level, income, physical activity, and diet, each of which has been suggested in previous research as potentially significant in influencing obesity rates (CDC, 2023). Pre-processing steps, such as standardizing column names, categorizing data where needed, and removing missing values, ensured data quality for accurate analysis.

The visualizations generated through bar charts provided an initial understanding of U.S. obesity patterns and are documented in Appendix A. These insights inform the following hypotheses and statistical tests.

Key Observations:

- **Obesity rates increase with Age:** As shown in Figure A1, obesity rates rise with age. Young adults (ages 18–24) exhibit the lowest obesity rates, while older adults (ages 45+) tend to have significantly higher rates.
- **Obesity rates decreases with Education:** Figure A2 shows a negative association between education level and obesity rates. Adults with college degrees or technical training exhibit lower obesity rates than those with a high school diploma or less.

- **Obesity decreases with Physical Activity:** Physical activity, such as aerobic and muscle-strengthening exercises, negatively correlates with obesity, as shown in Figure A3.

Hypotheses and Statistical Tests

Based on EDA observations, the following key questions were formulated and tested. The tests conducted are documented in Appendix B .

1. Hypotheses Test-1: Age and Obesity Rates

Statistical Question: As per the EDA, it was observed that the obesity increases with age. Is There a statistically significant difference in obesity rates across age groups?

- **Null Hypothesis (H0)**: The mean BMI of adults aged 18–24 years >= the mean BMI of adults aged 45–54 years.
- **Alternative Hypothesis (H1)**: The mean BMI of adults aged 18–24 years < that of adults aged 45–54 years

Test: A two-sample, left-tailed t-test was conducted on BMI means between younger (18–24) and older adults (45–54), as BMI is normally distributed, the two groups are independent, and variances were tested for equality with an F-test.

T-Test Result: From the Figure-B1, it can be observed that the T-test's P-value is less than commonly used significance level of 0.05, providing strong evidence to reject null hypothesis. Thus, we can conclude that obesity is more prevalent in older adults than younger ones.

Correlation Analysis: A correlation analysis (Figure-B2) with encoded dummy age categories showed a positive correlation (0.935), confirming that BMI increases with age.

2. Hypotheses Test-2 : Education Level and Obesity Rates

Statistical Question: Based on EDA, it was observed that the prevalence of adult obesity decreased as education level increased. Is there a statistically significant difference in obesity rates across educated and uneducated groups?

- **Null Hypothesis (H0)**: Mean BMI of non-graduates <= Graduates from College or Technical School
- **Alternative Hypothesis (H1)**: Mean BMI of non-graduates > Graduates from College or Technical School

Test: A **2-sample Left tailed T-test for BMI means** between graduate and non-graduates groups as the BMI is numerical, normal and the 2 groups selected are independent, has equal variance (as tested with F-test) is performed.

T-Test Result: From the Figure-B3, it can be observed that the T-test's P-value is less than commonly used significance level of 0.05, providing strong evidence to reject null hypothesis. Thus, we can validate that **obesity is more prevalent in non-graduates than graduates.**

Correlation Analysis: To test the direction of association, correlation test in Figure-B4 was executed on the encoded dummy categorical education variables, highest college graduate as '1' and less educated group, high school graduate as '4'. This provided a positive correlation value of 0.572 signifying that as the group increase BMI increases.

3. Hypotheses Test-3 : Physical Activity and Obesity

Statistical Question: With EDA we inferred that lack of Physical activity is a factor for obesity. This says that physical activity is a factor

- **Null Hypothesis(H0):** There is no or positive significant relationship between physical activity and BMI. (*Pearson's correlation >= 0)*
- **Alternative Hypothesis(H1):** There is a significant negative relationship between physical activity and BMI. (Pearson's correlation < 0)

Test: Pearson's correlation test was used to measure the association.

Test Results: The results (Figure B5) show a negative correlation of -0.1684, indicating that increased physical activity is associated with a reduction in obesity risk.

4. Hypotheses Test-4 : Regression Analysis:

Statistical Question: How do age, education, and physical activity predict BMI?

- **Null Hypothesis(H0): No statistically significant relationship** between the predictors and BMI classification, $\beta i=0$.
- **Null Hypothesis(H1): Statistically significant relationship** between the predictors and BMI classification βi not equals 0.

Test: Multiple regression analysis was conducted.

Test Results: All predictor coefficients were significantly different from zero ($p < 0.05$; see Figure B6), suggesting that age, education, and physical activity are valuable predictors of BMI.

Key Insights & Recommendations

The analysis provides a comprehensive look at factors influencing obesity in the U.S., highlighting:

1. **Age:** Obesity significantly increases with age, with older adults (aged 45-54) exhibiting higher mean BMI than younger adults (aged 18-24). Targeted interventions for older populations are essential.
2. **Education Level:** Higher obesity rates were observed among individuals without a college education. This implies, educational initiatives promoting healthier lifestyles could help reduce obesity in these groups.
3. **Physical Activity:** A strong negative correlation exists between physical activity levels and BMI, indicating that increased physical activity is crucial for obesity prevention.

Conclusion

The present analysis underscores that age, education level, and physical activity significantly influence obesity rates in the United States. Age emerged as a key predictor, with older adults showing higher obesity prevalence, consistent with national health statistics (Centers for Disease Control and Prevention [CDC], 2023). Educational attainment also correlated with obesity, as those without a college degree showed higher BMI levels, supporting the idea that educational interventions could play a role in promoting healthier lifestyles. Physical activity demonstrated a negative association with obesity, suggesting that it is an essential factor in managing body weight and reducing obesity-related health risks.

In conclusion, a holistic approach addressing both demographic and lifestyle factors is essential for combating obesity. These findings offer valuable insights for public health initiatives targeting age-specific and educationally-informed strategies to foster healthy habits and reduce the national burden of obesity.

Appendix A : Exploratory Data Analysis - Visualizations

This appendix consists of the visualizations generated during EDA of BRFSS dataset to understand obesity tends across demographics.

1. Obesity Distribution Across Age and Education

Figure A1 depicts that young adults(18-24) show the lowest obesity rates than elder adults. Figure A2 depicts that less educated adults(less than high school) show the highest obesity rates than college graduates.

Figure A1: Obesity Distribution Across Age Group.

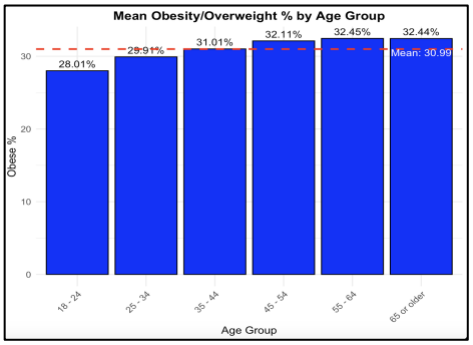
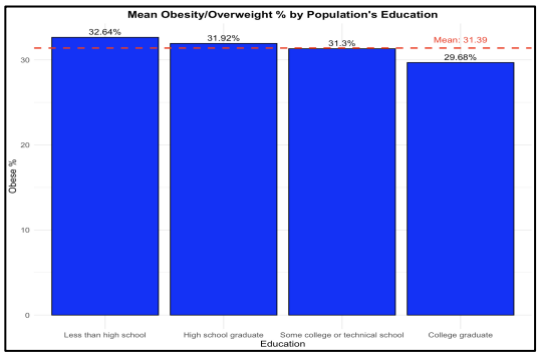


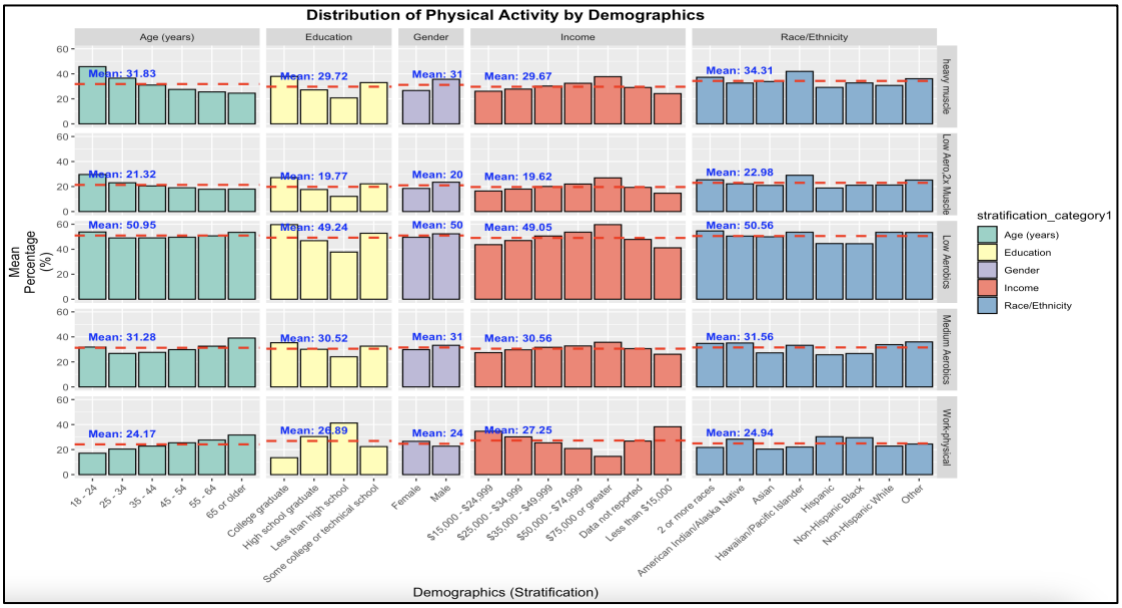
Figure A2: Obesity Distribution Across Education



2. Distribution of physical activity Across Demographics

Figure A3 show the obesity distribution of physical activity (muscle exercises, aerobic activity) across demographics.

Figure A3: Obesity across Demographics & Physical Activity



Appendix B : Hypothesis Test – Results

This appendix consists of the tables resulted during hypothesis tests of BRFSS dataset to understand obesity trends across demographics.

1. Hypothesis Test -1: Age Vs Obesity

Figure B1:Two-sample left tailed T-test proves Obesity increases with Age

	Statistic	Value
1	Tested Variable	Obesity Level Across Age
2	Null Hypothesis	H0: Obesity Level of adults aged 18-24 >= 45-54
3	Alternative Hypothesis	Ha: Obesity Level of adults aged 18-24 < 45-54
4	P-Value	< 2.22e-16
5	Significance Level	95% (0.05)
6	18_24_age_Group Mean	3.2297
7	45_54_age_Group	3.365
8	Result	Reject H0

Figure B2: Pearson's Test Shows Strong Positive Association (-0.9357)

Table: Pearson's Test to check Association between Age and Obesity/Overweight

Statistic	Value
Tested Variable	BMI And Age
Null Hypothesis	H0: Negative or No significant association
Alternative Hypothesis	Ha: Significant Positive association
Pearson's Coefficient	0.9357
Result	Reject H0, Significant Positive Association Exists

2. Hypothesis Test -2: Education Vs Obesity

Figure B3: Two-sample left tailed T-test proves Obesity increases with No Education

	Statistic	Value
1	Tested Variable	Normalized BMI Across Education Level
2	Null Hypothesis	H0: Mean BMI of No graduates <= Graduates
3	Alternative Hypothesis	Ha: Mean BMI of No graduate > Graduates
4	P-Value	< 2.22e-16
5	Significance Level	95% (0.05)
6	No_graduate_Group Mean	3.3504
7	Graduate_Group Mean	3.301
8	Result	Reject H0

Figure B4: Pearson's Test Shows Strong Positive Association (0.5721)

Table: Pearson's Test to check Association between Education and Obesity/Overweight

Statistic	Value
Tested Variable	BMI And Education (Encoded education college graduate to less than school, as 1 to 4)
Null Hypothesis	H0: Positive or No significant association
Alternative Hypothesis	Ha: Significant Negative association
Pearson's Coefficient	0.5721
Result	Reject H0, Significant Positive Association Exists

3. Hypothesis Test -3: Physical Activity Vs Obesity

Figure B5 : Pearson's Test Shows Association (-0.1684)

Table: Pearson's Test to check Association between Physical Activity and Obesity/Overweight

Statistic	Value
Tested Variable	Normalized BMI And Physical Activity
Null Hypothesis	H0: Positive or No significant association
Alternative Hypothesis	Ha: Significant Negative association
Pearson's Coefficient	-0.1684
Result	Reject H0, Significant Negative Association Exists

4. Regression Results:

Figure B6 : Regression coefficients for predictors and BMI.

Table: Logistic Regression Results for BMI Classification Model

	Estimate	Std. Error	t Value	p-value
(Intercept)	1.576	0.004	386.776	0
physical_act	0.038	0.001	53.646	0
educag	-0.019	0.001	-21.608	0
age_g	0.017	0.000	34.869	0

References:

Centers for Disease Control and Prevention. (2021). **Adult obesity facts**. https://www.cdc.gov/obesity/php/data-research/adult-obesity-facts.html

Centers for Disease Control and Prevention. (2023). **Behavioral risk factor surveillance system: 2023 survey data and documentation**. https://www.cdc.gov/brfss/annual_data/annual_2023.html

Centers for Disease Control and Prevention. (2021). **Behavioral risk factor surveillance system: 2021 survey data and documentation**. https://www.cdc.gov/brfss/annual_data/annual_2021.html

Centers for Disease Control and Prevention. (2023). **Adult obesity prevalence maps**. https://www.cdc.gov/obesity/php/data-research/adult-obesity-prevalence-maps.html

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). (n.d.). **Overweight and obesity statistics**. https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity

U.S. Department of Agriculture Economic Research Service. (2023, November). **Researchers adjust self-reported estimates of obesity in scanner data**. https://www.ers.usda.gov/amber-waves/2023/november/researchers-adjust-self-reported-estimates-of-obesity-in-scanner-data/

Centers for Disease Control and Prevention. (2023). **2023 BRFSS codebook**. https://www.cdc.gov/brfss/annual_data/2023/zip/codebook23_llcp-v2-508.zip

National Center for Biotechnology Information (NCBI). (n.d.). **Significance level in hypothesis testing**. https://www.ncbi.nlm.nih.gov/

Centers for Disease Control and Prevention. (n.d.). **Nutrition, Physical Activity, and Obesity Behavioral Risk Factor Surveillance System (BRFSS)**. Chronic Data. https://chronicdata.cdc.gov/Nutrition-Physical-Activity-and-Obesity/Nutrition-Physical-Activity-and-Obesity-Behavioral/hn4x-zwk7 (Website as of October 3)

U.S. Government Open Data Portal. (n.d.). **Nutrition, Physical Activity, and Obesity Behavioral Risk Factor Surveillance System**. Data.gov. https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system (Website as of October 3)