Adult Obesity Rates and Its Impact in the United States

Ayushi Yadav, Monika Kommareddy, Pratik Pachpute, Zhenjun Wang

Abstract This analysis of the Adult Obesity Rate in America provides a complete overview of the current state of obesity in the United States, including time trends, disparities, and other factors. BMI (weight (kg)/height (m)2) cut points of 25 and 30 were used to classify adults as overweight and obese, respectively. The goal of the report is to investigate how environmental variables influence obesity. The adult obesity rate is used as the dependent variable (Y) while all other parameters such as physical activity, number of families with a car, household income, and so on are used as the independent variables (X). Obesity is influenced by a variety of factors, including the cost of packaged snacks against the cost of fresh vegetables.

Obesity rates in a substantial portion of the country are increasing by the day. Some states are more affected than others. Texas appears to be the most affected by all the factors we used in the study. Obesity affects more than 35% of the population today, and the future prevalence is expected to rise.

Key Words: Obesity, Poverty Rate, United States of America, Adult Obesity,
Topic: Obesity, Adult Obesity, Percentage of Adults exercising, Percentage of Smokers

1. Introduction

According to national surveys, the prevalence of obesity has risen and continues to rise in the United States over the last two decades. Obesity affects more than one-third of the population in the United States. Obesity is associated with an increased risk of heart disease, type 2 diabetes (high blood sugar), high blood pressure, certain cancers, and other chronic conditions, in addition to girth. Obesity has become a complex problem in recent years, with multiple factors influencing the adult obesity rate.

Environmental factors or economic situations can have a significant impact on rising obesity. The current report only includes factors in the American environment that indirectly contribute to the rising obesity rate. The rising number of obese people in the country drives up healthcare costs dramatically. The cost of direct and indirect health care in the United States is estimated to be between $147 billion and $210 billion. However, the financial aspect is not the only concern. Obesity can cause premature death and increase susceptibility to other severe chronic diseases, as well as have an incalculable impact on the quality of life and the family.

Knowing the factors that contribute to obesity, either directly or indirectly, can aid in the development of prevention strategies that foster healthy lifestyles.

2. Project Methodology

2.1. Exploratory Data Analysis

We approached the data by observing the general obesity trends in the country. (Fig.1). It was observed that obesity rates are higher in the eastern and central regions of the country when compared to their western counterparts.

2.2. Statistical Model and Inference

Data Screening – The first step of the analysis was to clean and prepare the data. We checked the data for Accuracy, Missing values, Outliers and Assumptions. The data appeared to be accurate with no missing values. An

initial regression was performed to detect the outliers on the dataset inclusive of the 6 variables used. We used Leverage, Cooks, and Mahalanobis to detect and exclude the outliers. There were 3 common outliers common to all the 3 measures used, these observations were excluded from the dataset.

We then used the no_outliers dataset to perform an initial regression to test the data for assumptions. We conducted tests for Additivity, Linearity, Normality, Homogeneity, and Homoscedasticity. As seen in figures 3-7 below, we can say that the data successfully met all assumptions.

The correlation plot suggested that 6 out of the 13 variables had a strong correlation of 0.6 with our response variable, obesity. Household income was highly positively correlated with the other predictor Poverty rate and was excluded from the dataset for further analysis.

2.3. "Best" Model

We then used the final_noout dataset excluding the variable household Income, to avoid Multicollinearity, and then performed real Regression on this dataset, termed as a model. After adjusting for correlated variables, the final model included just 5 of the 22 variables from the original dataset. With an R-Squared value of 0.84, our model has a very strong fit and predicts the data 84% of the time. A description of our significant variable and model statistics is given below.

Smokers – Yes, b = 0.167, t (43) = -2.627, p = 0.011

no_car – No, b = -0.66, t (43) = -1.77 (small), p = 0.08(greater than0.05)

No_exercise – Yes, b = 1.78, t (43) = 6.148, p = 0.00000023

Gals_soft_drinks – No, b= 0.09, t (43) = 1.653 (small), p = 0.1055(greater than 0.05)

Poverty_rate – No, b = -0.171, t (43) = -1.208 (small), p = 0.233(greater than 0.05)

We built 2 more models in addition to the one above, with different predictor variables, the summary of which is included in Table 1. On comparing the R-squared values for all three models, it appeared that model1 with all the 5 independent variables had the highest R-squared value and was considered the best fit model. It was interesting though, to see that model 3 had the minimum Root Mean Squared Error value out of all three models and model1 on the other hand had the highest R-squared value.

Regression Equation:

Obesity = -19.804 + 0.208*Smokers – 0.466*No_car + 1.78*No_exercise + 0.100*Gals_soft_drinks - 0.345*Poverty_rate

2.4. Prediction

We created a training dataset to train the chosen model and a separate test dataset with a 5% increase in our significant variables to test the model. As can be seen in Fig. 7 and Fig. 8 the predicted data is very close to the actual data.

The accuracy of our prediction is explained by the low Mean Absolute Percentage Error and Root Mean Square Error-values which were minimal of the 3 models and in the acceptable range.

MAPE – 0.049
RMSE – 1.67

3. Conclusion

It was observed that the Not Exercising predictor had a significant Linear positive impact on the obesity rates. In addition, Smoking and Obesity also displayed a significant positive relationship. Poverty was

the only negative significant predictor, and the no car variable was the only non-significant predictor from the variables considered for our analysis. To summarize, obesity has affected many major regions of the United States and is growing at an alarming rate. The relationship between obesity and household income, recreational facilities available, and the percentage of adults exercising are complex and dynamic. Considering all these facts, some population-based awareness programs and strategies are required to be developed.

4. Limitations and Caveats

We are operating under the assumption that averages are approximately equal for all counties and the difference in urban and rural setting rates is negligible.

A more in-depth analysis needs to be conducted which includes variables like healthcare facilities, education, gender, ethnicity, and so on.

5. Figures and Tables

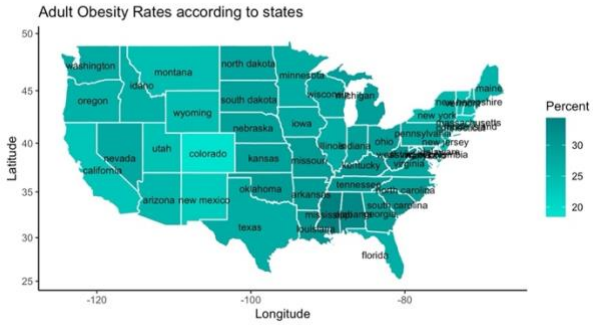

Fig 1. *General Obesity Trends in the United States*

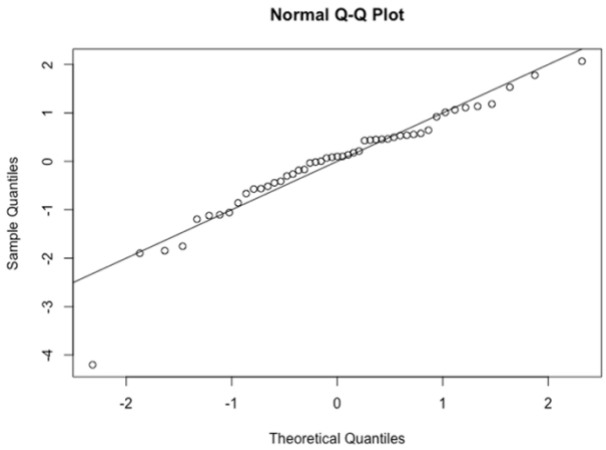

Fig 2. *Normal Q-Q plot – Checking Linearity of data*

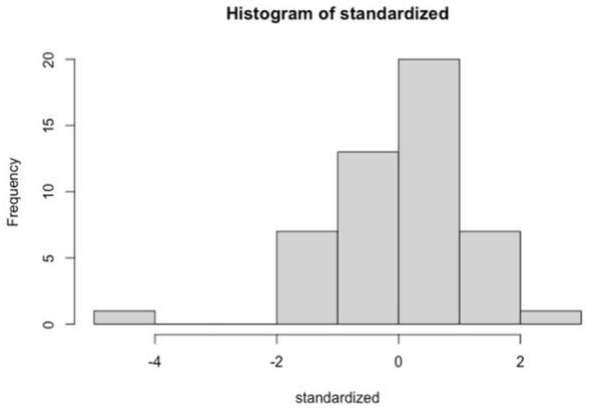

Fig 3. *Histogram of fitted Values – Checking Normality of data*

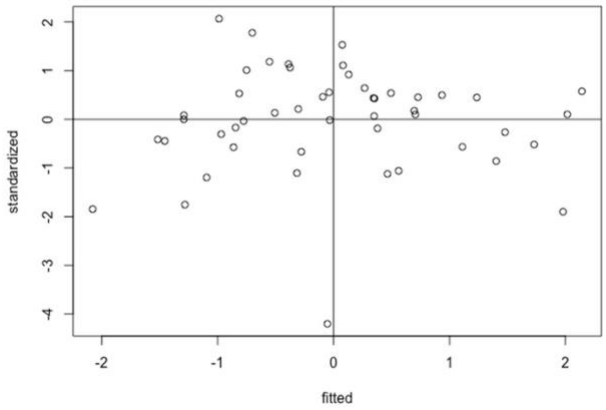

Fig 4. *Homogeneity and Homoscedasticity check*

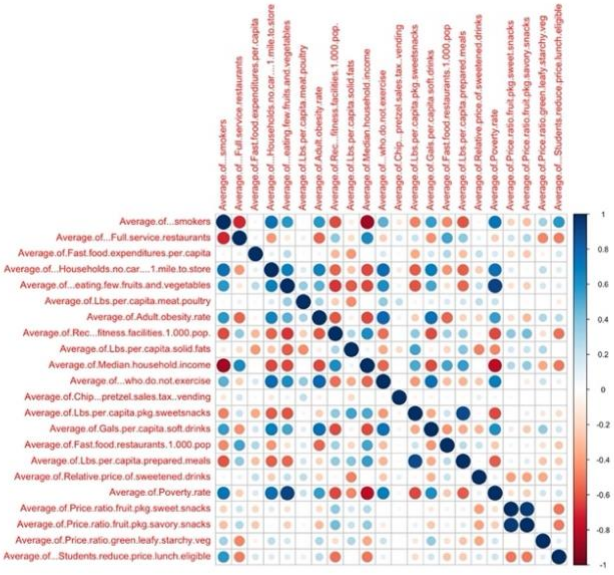

Fig 5. *Correlation Matrix Plot– Checking Correlation between Independent Variables.*

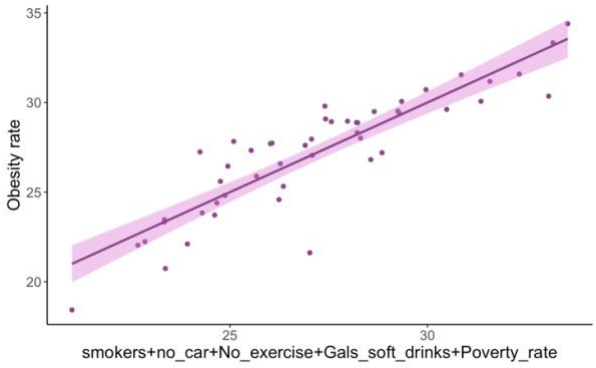

Fig 6. *Correlation Matrix Plot– Checking Correlation between Independent Variables.*

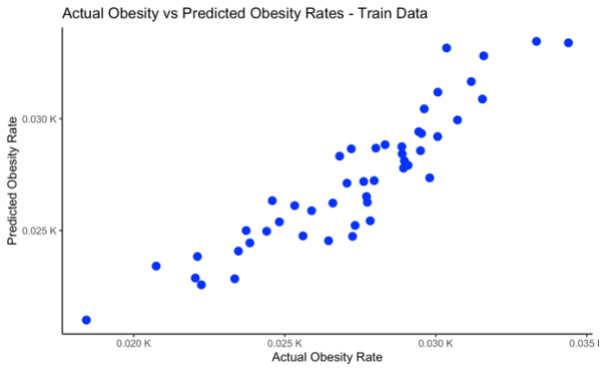

Fig 7. *Training the model*

Fig 8. *Prediction with the final model*

Table 1. *Summary of Models*

Model	Variables Considered	R-2	Prediction Accuracy	Comments
1.	Smokers no_car No_exercise Gals_soft_drinks Poverty_rate	Multiple – 0.8428 Adjusted – 0.8241	MAPE – 0.049 RMSE – 1.67	
2	Poverty Rate	Multiple – 0.369 Adjusted – 0.3563	MAPE – 0.091 RMSE – 2.91	Only Poverty Rate
3.	Smokers Poverty Rate No_exercise Gals_soft_drinks	Multiple – 0.8349 Adjusted – 0.8196	MAPE – 0.055 RMSE – 1.6	Dropped no_car