NORTHERN ILLINOIS UNIVERSITY

Exploring the Spatial Association between Obesity and Selected Underlying factors among Adults in the United States

A Capstone Submitted to the
University Honors Program
In Partial Fulfillment of the
Requirements of the Baccalaureate Degree
With Honors
Department Of
Earth, Atmosphere and Environment

By
Xinyi Ye
DeKalb, Illinois
May, 2022
University Honors Program
Capstone Faculty Approval Page

Capstone Title (print or type)
Exploring the Spatial Association between Obesity and Selected Underlying factors among Adults in the United States

Student Name (print or type) ________________ Xinyi Ye ________________

Faculty Supervisor (print or type) ________________ Xuwei Chen ________________

Faculty Approval Signature ________________

Department of (print or type) _______ Earth, Atmosphere and Environment _______

Date of Approval (print or type) ________________ 04/27/2022 ________________

Date and Venue of Presentation _______ CURE _______ 04/26/2022 _______

Check if any of the following apply, and please tell us where and how it was published:

  Capstone has been published (Journal/Outlet):

  ________________

  Capstone has been submitted for publication (Journal/Outlet):

  ________________

Completed Honors Capstone projects may be used for student reference purposes, both electronically and in the Honors Capstone Library (CLB 110).

If you would like to opt out and not have this student's completed capstone used for reference purposes, please initial here: _______ (Faculty Supervisor)
HONORS CAPSTONE ABSTRACT

According to CDC, obesity and diseases caused by obesity in the United States have become critical issues in the population. Existing literature has identified a number of factors that play roles in the progression and development of obesity. However, these studies tend to focus on smaller geographic regions and single factors in obesity. Under this context, this research aims to investigate the obesity pattern and its associated factors at the national level from a spatial perspective. Based on 2017 obesity rates at the county level, this study first explores the overall spatial correlations of obesity using Global Moran’s I and the geographic clusters of high/low obesity rates in the contiguous US. Furthermore, this study employs OLS to identify what socio-economic factors and health indicators are statistically related to obesity in general, then applies the spatial error and geographic regression models to reveal the spatial associations between obesity and those factors.

The findings revealed a clear spatial pattern of obesity in this country. Higher obesity rates were observed in the southeast, especially in Mississippi and the neighboring states. States in the west tended to have lower obesity rates. Local hot spots were revealed predominantly in eastern Mississippi, South Carolina, and West Virginia whereas clusters of low obesity rates were primarily observed in the West, around New Jersey, some parts of Texas and southern Florida.
Exploring the Spatial Association between Obesity and Selected Underlying factors among Adults in the United States

1. Introduction

According to the World Health Organization, the rate of increase in global obesity in recent years is a big concern. Since 1975, the global prevalence of obesity has almost tripled, more people are overweight or obese today than at any other time in history. (World Health Organization, 2021) The United States is one of the most obese regions, according to the Centers for Disease Control and Prevention (CDC), the prevalence of obesity in the United States has increased from 30.5% to 42.4%, which has the 12th highest obesity rate in the world. Although this ranking doesn't seem very high, it is worth noting that obesity rates vary widely from state to state in the US due to dietary, social, economic, environmental, and cultural factors. (World Health Organization, 2022)

According to the report from the most recent Behavioral Risk Factor Detection System (BRFSS), 16 states have adult obesity rates above 35%. (Robert Wood Johnson Foundation, 2020) Epidemiological studies have shown a substantial increase in overweight and obesity-related mortality, with obese people having a 50% to 100% increased risk of premature death from all causes. It is estimated that around 300,000 deaths each year can be attributed to obesity. (U.S. Department of Health and Human Services, 2010)

2. Literature Review

The role of various factors in the formation and development of obesity is well established, with genetic, behavioral, and environmental factors being important factors associated with obesity. Studies have examined the association between these various potential factors and obesity. Morland et al. provide evidence that high supermarket accessibility has a negative impact on obesity and high fast food restaurant accessibility has a positive impact on obesity. (Morland, 2009) Huang et al. found an inverse association between convenience stores and obesity while demonstrating that race is a
moderator of the relationship between food environment and obesity.(Huang, 2021) Casey et al. demonstrated that the socioeconomic status of individuals influences the relationship between spatial accessibility of public facilities and obesity.(Casey, 2012) Meanwhile, Wang et al. found that race and socioeconomic status may contribute to differences in diet, exercise, and weight status in the United States.(Wang, 2011) Ellaway et al. found that adults' BMI was associated with different modes of transportation, especially walking and biking.

Measures the accessibility of public broadcasting facilities. Most studies report that behaviors affecting obesity include unhealthy lifestyles (consumption of unhealthy foods, low physical activity, etc.); environmental factors include a long-term presence in poor neighborhoods, education, and lower-income communities.(Zgodic, 2021) It is worth noting that one study suggests that obesity risk may be related to the political orientation of the county in which one lives.(McCarthy, 2013)

Since the beginning of the new century, the prevalence of obesity in the United States has risen to 42.4%. Obesity has been a major problem in the U.S. According to numerous reports and studies related to the obesity level in the U.S. at the state level, it is clear that the U.S. obesity landscape presents a very different picture. It is likely that the variation of obesity rates across the states are associated with the variation of those different factors.(Lee, 2019) While numerous reports have revealed strong interactions between obesity status and individual correlated influencers within the same space, few studies have examined the interactions and effects of multiple factors simultaneously on obesity in U.S. adults. Meanwhile, many studies prefer to identify clusters in smaller regions (community level or county level) by using Moran's I statistic and Local Indicator of Spatial Association (LISA) or GetisOrdGi* statistic.(Pouliou, 2009 and Penney, 2013) On the basis of spatial autocorrelation, some researchers often use regression models such as OLS or GWR to conduct multi-scale spatial analyses of obesity determinants. However, most of these studies focused on the spatial distribution at the same time and did not provide a deep analysis of the dynamic development of obesity and its influencing factors. This research aims to investigate the obesity pattern and its associated factors at the national level from a spatial perspective.
3. Study Area and Data

This study focuses on obesity at the national level in the United States and other potential factors that may influence obesity status. In order to exclude errors due to spatial blockage, this study selects the contiguous continental United States as the study area to facilitate the observation of spatial distribution patterns of obesity.

This study chose the county level as the smallest unit of observation for obesity and its influencing factors because state policies can vary greatly from state to state. We needed more detailed data within each state in order to exclude, to some extent, the influence of policies, the census track and more detailed data are difficult to obtain, and because the total population is small, its fluctuations are greater, and it does not accurately reflect the macroscopic situation of adult obesity and the distribution of factors affecting obesity.

We selected data provided by the University of Wisconsin Population Health Institute, which contains health data and underlying data at all county levels in the United States. Through literature reading and data interpretation, we selected factors that have a high probability of influencing obesity status, such as personal health behavior, health facility level, and medical condition status, and then examined them in more depth, relation to the underlying context of each county (population age, race, economic status, etc.) in an attempt to discover their precise association with adult obesity status.

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>sd</th>
<th>95% conf.</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physically Inactive</td>
<td>0.09</td>
<td>0.3</td>
<td>0.287795301</td>
<td>0.05648883</td>
<td>0.0021</td>
<td>0.454</td>
</tr>
<tr>
<td>Exercise Access</td>
<td>0</td>
<td>1</td>
<td>0.624856775</td>
<td>0.232163812</td>
<td>0.0083</td>
<td>-0.253</td>
</tr>
<tr>
<td>Healthy Food Access</td>
<td>0</td>
<td>0.72</td>
<td>0.085521373</td>
<td>0.082310311</td>
<td>0.0013</td>
<td>0.003</td>
</tr>
<tr>
<td>Food Insecure</td>
<td>0.04</td>
<td>0.3</td>
<td>0.132318887</td>
<td>0.03764109</td>
<td>0.0294</td>
<td>0.344</td>
</tr>
<tr>
<td>Mentally Unhealthy</td>
<td>2.7</td>
<td>7.3</td>
<td>4.67906257</td>
<td>0.66379865</td>
<td>0.0237</td>
<td>0.345</td>
</tr>
<tr>
<td>Physically Unhealthy</td>
<td>2.4</td>
<td>7.2</td>
<td>4.3918548</td>
<td>0.77225443</td>
<td>0.0276</td>
<td>0.356</td>
</tr>
<tr>
<td>Insufficient Sleep</td>
<td>0.26</td>
<td>0.49</td>
<td>0.368278052</td>
<td>0.039747707</td>
<td>0.0014</td>
<td>0.369</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.007</td>
<td>0.183</td>
<td>0.040143547</td>
<td>0.014206917</td>
<td>0.0005</td>
<td>0.252</td>
</tr>
<tr>
<td>Elder</td>
<td>0.049</td>
<td>0.582</td>
<td>0.198437078</td>
<td>0.047571153</td>
<td>0.0017</td>
<td>-0.087</td>
</tr>
<tr>
<td>Black</td>
<td>0</td>
<td>0.859</td>
<td>0.091520438</td>
<td>0.143256657</td>
<td>0.0051</td>
<td>0.297</td>
</tr>
<tr>
<td>Uninsured</td>
<td>0.007</td>
<td>0.43</td>
<td>0.135117477</td>
<td>0.063013216</td>
<td>0.0022</td>
<td>0.052</td>
</tr>
<tr>
<td>Median Household</td>
<td>$24,700</td>
<td>$151,800</td>
<td>$95,365</td>
<td>$14,427</td>
<td>$315</td>
<td>$315</td>
</tr>
<tr>
<td>Adults Obesity</td>
<td>0.11</td>
<td>0.59</td>
<td>0.335041854</td>
<td>0.056583888</td>
<td>0.0021</td>
<td></td>
</tr>
</tbody>
</table>

4. Methodology
The whole research can be roughly divided into two parts, the first part is to explore the spatial pattern of obesity by studying the spatial autocorrelation. In the Global Moran’s I test, we test the entire study area as a whole to obtain an overall trend for a variable. Thus, in this section we use this model to explore the overall spatial associations.

\[ I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}) (\sum_{i=1}^{n} (x_i - \bar{x}))} \]

In this formula, \( n \) is the number of the counties. \( x \) is the value of a specific independent variable in that county, which means, in this study, \( x \) shows the obesity rate of adults in a specific county. And \( w_{ij} \) is the weight setting between county \( i \) and \( j \).

In the second section, we use Local Moran’s I(LISA) to identify areas with high or low obesity rates. In this formula, \( j \) is the neighbor of county \( i \) and \( z \) shows the \( z \) score of a specific county.

\[ I_i = z_i \sum_j w_{ij} z_j \]

After finding out the spatial pattern of obesity, we use regression models to identify the potential factors. First we use the Ordinary Least Squares (OLS) Model, in this formula, \( y \) shows the \( i^{th} \) observation of the dependent variable, \( \beta \) is the coefficient of a specific independent variable and \( \varepsilon \) is the error coefficient.

In this study, we chose adult obesity rate as the dependent variable and chose physically inactive rate, exercise access rate, healthy food access rate, food insecure, insufficient sleep rate, unemployed rate, elder rate, black rate and uninsured rate as independent variables.

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \]
Although the OLS model is very effective and simple in statistics, because it is not a spatial regression model, it can only perform statistical processing on the selected data and ignoring the influence on the geographical ways. Thus, in order to find out more accurate results, we use the Spatial Error Model and the Geographically Weighted Regression (GWR) Model to process the data.

\[ y = X \beta + \lambda W u + \varepsilon \]
\[ y(g) = \beta_0(g) + \sum_{k=1}^{m} \beta_k(g) x_k + \varepsilon \]

The figure on the upper left is the formula of the Spatial Error Model, \( \lambda \) shows the coefficient of weighted errors and \( W \) is the errors for weight matrix. The figure on the upper right is the formula of Geographically Weighted Regression Model, \( \beta(g) \) is the parameters to be estimated at a location whose coordinates are given by the vector \( g \).

5. Results

I first examined obesity in the United States, and found out just like the result from previous research, higher obesity rates were observed in the southeast, especially in Mississippi and the neighboring states. States in the west tended to have lower obesity rates.
After finding out the obesity rate distribution, I analyze the results of spatial autocorrelation detection and try to find out the obesity spatial pattern. The figure below shows the overall distribution trend of adult obesity rate in the entire research field, the result of Global Moren's I test is 0.312 (positive), which shows that the adult obesity rate has a certain clustered trend in the United States, but this trend is not very strong.

Local Moren's I test analyzes the correlation between the county and its neighbors, when a county and its neighboring counties have high obesity rates, they are shown in red in the graph, reflecting hotspots; and when a county and its surrounding counties have low obesity rates, they are shown in blue in the figure, it is reflected in the cold spot area. In this Local Moren's I test (upper right figure), we can clearly find that obesity hotspots are concentrated in the eastern part of the United States, around Mississippi, South Carolina and West Virginia. The obesity cold spots are mainly concentrated in the western United States, near New Jersey, some parts of Texas and southern Florida.

In the regression study, using the OLS model, I found out that not all independent variables are significant. After removing insignificant variables, I got the following table. From this table, we can find that Physically Inactive rate, Insufficient Sleep rate and Unemployed rate play the biggest role in the distribution of obesity in the OLS model.
The figure below shows the predicted value of the OLS model for the Obesity situation. We can clearly find that the U-shaped region in the southeast has a higher predicted value. Since the predicted value can show the difference between the model results and the real situation, it is explained that the place with high predicted value in the model has worse obesity situation in the real world. According to this, we can say obesity around Mississippi and around Northern Virginia was much worse than the model expected. Interestingly, these two places are also hotspots for obesity.

To further optimize the model, I used the Spatial Error Model and Geographically Weighted
Regression Model on the same variables. We can find that the following table does not have a huge change from the OLS table except for some coefficient changes, which also proves that these potential factors do have an impact on adult obesity.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>0.1039</td>
<td>0.0102</td>
<td>10.1972</td>
<td>0</td>
</tr>
<tr>
<td>Physically Inactive</td>
<td>0.3550</td>
<td>0.0201</td>
<td>17.6515</td>
<td>0</td>
</tr>
<tr>
<td>Insufficient Sleep</td>
<td>0.3866</td>
<td>0.0326</td>
<td>11.8474</td>
<td>0</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.3671</td>
<td>0.0834</td>
<td>4.4020</td>
<td>0</td>
</tr>
<tr>
<td>Elder</td>
<td>-0.0883</td>
<td>0.0210</td>
<td>-4.2024</td>
<td>0</td>
</tr>
<tr>
<td>Black</td>
<td>0.0153</td>
<td>0.0096</td>
<td>1.5961</td>
<td>0.1105</td>
</tr>
<tr>
<td>Uninsured</td>
<td>-0.0409</td>
<td>0.0208</td>
<td>-1.9657</td>
<td>0.0493</td>
</tr>
<tr>
<td>LAMBDA</td>
<td>0.3961</td>
<td>0.0241</td>
<td>16.4162</td>
<td>0</td>
</tr>
</tbody>
</table>

However, through the above figure, we can clearly find that the part with high Predicted value has become less, and it can be clearly found that the Spatial Error Model can more accurately obtain
the real situation of the obesity rate. However, we can still observe the high predicted value for Northern Virginia in the plot, and the "edges" that "connect" between several high predicted values.

By looking at the Predicted Obesity Rates of the GWR model, we were able to find that obesity rates in general were greatly improved, especially in Northern Virginia and on both sides of the U-shaped region. But Mississippi's underestimation of obesity rates has not improved very well, and there have been cases where individual counties have severely underestimated obesity rates. Also of concern is the overestimation of obesity rates in the GWR model, especially on the west coast of the United States and in Colorado, more so than in the previous models.

Nonetheless, in this study we will focus more on the situation in areas with high obesity rates, so the GWR model has a greater advantage than the first two models. AIC (AICc) and R square are parameters used to compare the accuracy of the model. The larger the absolute value of the two, the
more accurate the model. From the table below, we can also see that the GWR model has the highest accuracy and is the most suitable model in this study.

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>AlCc</th>
<th>R Square</th>
<th>Adjusted R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS Model</td>
<td>-9653.32</td>
<td>0.3354</td>
<td></td>
<td>0.334090</td>
</tr>
<tr>
<td>Spatial Error Model</td>
<td>-9889.26</td>
<td>0.4025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWR Model</td>
<td>-10272.8681</td>
<td>0.5290</td>
<td>0.4775</td>
<td></td>
</tr>
</tbody>
</table>

After looking at how all independent variables affect adult obesity rates together, I found out the GWR model suggests that there were varying effects of those underlying factors on obesity geographically. The figure below shows the coefficients of physically inactive and insufficient sleep, respectively. We can obviously found out the physically inactive was a more influential factor in the west and northeast than in other regions whereas it was insufficient sleep in some southern states, e.g. Texas and Louisiana.

6. Conclusions

All in all we can conclude that higher obesity rates were observed in the southeast, especially in Mississippi and the neighboring states. States in the west tended to have lower obesity rates. There was a clear spatial pattern of obesity in the country. Local hot spots were revealed predominantly in eastern Mississippi, South Carolina, and West Virginia whereas clusters of low obesity rates were
primarily observed in the West, around New Jersey, some parts of Texas and southern Florida. Unemployed rate, physically inactive rate and insufficient sleep rate showed stronger associations with adult obesity rate at the county level. And these influencing factors have varying effects on obesity geographically.

7. Weaknesses and Discussion

This study did not examine the development of obesity over time in detail due to time issues, and did not analyze emerging hotspots or coldspots using emerging hotspot analysis techniques. This makes this study of obesity in the United States not rich enough.

At the same time, during the research I found that although food index and opportunities to access exercise locations have been mentioned in many studies as important indicators affecting obesity. This study did not identify a significant correlation between them. This is caused by the fact that these two factors are strongly influenced by accessibility, which usually has a stronger effect on smaller geographic units (e.g. community).

8. References


Wang Y, Chen X: How much of racial/ethnic disparities in dietary intakes, exercise, and weight status can be explained by nutritionand health-related psychosocial factors and socioeconomic status among US adults?. J Am Diet Assoc 2011;111:1904-1911


