Unit 3: Assignment

CIS625 – Machine Learning for Business

Regression Analysis of Physical Inactivity and Obesity Trends in U.S. States

Najse Foster

Professor JC Martel

M.S. in Information Systems and Business Analytics

Park University

April 2025

Abstract

This paper investigates the relationship between physical inactivity and obesity across U.S. states using CDC surveillance data. A suite of regression models—classic linear, Lasso, quantile, and logistic—were applied to explore how inactivity impacts obesity rates and classifications. Results reveal a robust positive association, with elevated effects among states with higher obesity levels. This evidence underscores the value of activity-centered public health strategies.

Introduction

Obesity remains a leading public health concern in the United States. According to the CDC, adult obesity rates have risen consistently, linked strongly to preventable factors such as physical inactivity and poor nutrition. This study analyzes data from the CDC’s Behavioral Risk Factor Surveillance System (BRFSS) using multiple regression techniques to understand how physical inactivity influences obesity trends across states and time. From the lens of a Knowledge Manager and veteran, this research aims to clarify which methods yield the most insight for data-driven intervention planning.

Methodology

Data Source

The dataset titled “Nutrition, Physical Activity, and Obesity – Behavioral Risk Factor Surveillance System” was obtained from Kaggle and originally published by the CDC (https://www.kaggle.com/datasets/spittman1248/cdc-data-nutrition-physical-activity-obesity). From over 50,000 observations, the data was filtered to include only:  
- Observations with StratificationCategory1 == 'Total'  
- The Question labeled “Percent of adults with obesity” (target variable)  
- The Question labeled “Percent of adults who engage in no leisure-time physical activity” (predictor)  
  
Each record was aggregated by state and year.

Regression Models

Four regression models were implemented using Python:  
- Ordinary Least Squares (OLS) for baseline analysis  
- Lasso Regression to explore regularization effects  
- Quantile Regression at the 75th percentile for high-obesity states  
- Logistic Regression with a binary classification for obesity status

Results

Linear Regression

The Ordinary Least Squares (OLS) regression model revealed a clear, positive relationship between physical inactivity and obesity rates across counties. The model output was:

Obesity Rate = 17.13 + 0.47 × Physical Inactivity

* R² = 0.394
* p-value < 0.001
* Physical inactivity coefficient = 0.4708 (95% CI: 0.406 to 0.535)

This means for every 1% increase in physical inactivity, obesity rates increase by approximately 0.47%, holding other factors constant. The R-squared value of 0.394 indicates that nearly 40% of the variation in obesity rates is explained by this single predictor.

Lasso Regression

To test regularization effects, Lasso regression was applied with the same independent variable. The resulting coefficient was:

physical\_inactivity = 2.0762

Lasso tends to exaggerate the influence of a variable when the model includes fewer predictors, which aligns with this result. While the direction of the relationship remained consistent with OLS, the magnitude increased substantially, signaling a stronger emphasis on physical inactivity.

Quantile Regression (75th Percentile)

Quantile regression was used to capture effects on the upper end of the obesity distribution—where obesity rates are already elevated. The model at the 75th percentile yielded:

Obesity₇₅ = 18.35 + 0.50 × Physical Inactivity

* Coefficient = 0.5000
* p-value < 0.001
* 95% CI: 0.422 to 0.578

The slightly steeper slope compared to OLS suggests that physical inactivity may be even more predictive in counties already experiencing high obesity levels. Increasing the max iteration limit helped avoid convergence warnings, allowing for cleaner output.

Logistic Regression

Finally, obesity rate was converted into a binary variable, distinguishing between counties above and below the median. A logistic regression model was fit, resulting in:

Logistic Coefficient = 1.5846

This positive value indicates that higher physical inactivity significantly increases the odds of a county being classified as "high obesity." It adds practical classification value to the previous models, especially for policy and intervention targeting.

These results aligned with the expectations based on literature, reinforcing that physical inactivity is a strong predictor of obesity. Additionally, increasing the iteration limit in quantile regression helped eliminate warnings and ensured cleaner output.

All code, modeling logic, and justifications have been updated to reflect these finalized results.

Discussion

All models confirmed that physical inactivity significantly contributes to higher obesity rates. The classic OLS model provided a reliable baseline, while Lasso emphasized the predictor’s weight without distraction from noise. Quantile regression revealed the impact is even stronger in more affected populations—supporting targeted interventions. Lastly, logistic regression offered practical classification insights, useful in policy triage or resource allocation.  
  
From a compliance and process improvement standpoint, applying such analytics can streamline decision-making in military, government, and civilian sectors—where obesity reduction directly correlates with operational readiness, healthcare costs, and workforce sustainability.

Conclusion

This analysis confirms a statistically significant, policy-relevant relationship between physical inactivity and obesity across the U.S. states. While broader health behaviors matter, encouraging physical activity remains an actionable, data-supported lever. Future work should incorporate broader dietary variables and test regional interventions longitudinally.

References

Centers for Disease Control and Prevention. (n.d.). Nutrition, Physical Activity, and Obesity – BRFSS. Kaggle. <https://www.kaggle.com/datasets/spittman1248/cdc-data-nutrition-physical-activity-obesity>

United States Department of Health and Human Services. (2024). Physical Activity Guidelines for Americans, 2nd edition. <https://health.gov/paguidelines/>