Physical Inactivity and Adult Obesity in US Counties*

Abhiram Bhokre

September 24, 2025

In this paper, I examine whether county-level physical inactivity is linked to adult obesity across U.S. counties. Using the 2023 County Health Rankings & Roadmaps dataset(County Health Rankings & Roadmaps 2023), I fit a simple linear regression with adult obesity (percent of adults with BMI 30) as the outcome and physical inactivity (percent of adults reporting no leisure-time physical activity) as the predictor. The results show a strong positive relationship: each one-percentage-point increase in inactivity is associated with an estimated 0.71 percentage-point increase in adult obesity (95% confidence interval: 0.70 to 0.73). The model explains about 63% of the variation in obesity across counties (n = 3,191). These findings are ecological and descriptive; they highlight a robust county-level association but do not imply individual-level causation.

Introduction

Obesity remains a prominent public health challenge in the United States and is closely linked to chronic conditions such as type 2 diabetes, cardiovascular disease, and certain cancers. At the same time, many communities report substantial levels of physical inactivity—adults who do not engage in leisure-time physical activity. Because both outcomes and behaviors vary widely across places, county-level comparisons offer a useful lens for understanding how community circumstances relate to health.

This project investigates whether counties with higher physical inactivity also tend to have higher adult obesity. I focus on the 2023 County Health Rankings & Roadmaps (CHR&R) dataset, which provides comparable, county-level indicators assembled from established surveys and administrative sources. Two measures are central here: (1) adult obesity (%), the share of adults with BMI 30, and (2) physical inactivity (%), the share of adults reporting

*Project repository available at: https://github.com/abhirambhokre1408/Math_261A_Paper_1.

no leisure-time physical activity. These indicators are widely used in community health assessments and grant applications, making a simple, transparent analysis especially relevant for local decision-makers.

I am fitting a simple linear regression with adult obesity as the outcome and physical inactivity as the predictor, using the most recent available year. This approach yields an interpretable summary of the association—how much obesity tends to increase, on average, as inactivity rises by one percentage point—while keeping the modeling assumptions and diagnostics accessible to a broad audience. The analysis is ecological and cross-sectional: it describes place-level patterns rather than individual behavior, and it does not establish causal effects.

The primary research questions are:

1. Do U.S. counties with higher physical inactivity also have higher adult obesity?
2. How large is the average change in obesity associated with a one-percentage-point increase in inactivity?
3. How much of the between-county variation in obesity can be summarized by a single linear predictor—physical inactivity?

This study makes three contributions. First, it provides a reproducible pipeline—from raw CHR&R spreadsheets to a cleaned analysis file—so that results can be regenerated or extended to future years. Second, it presents a clear, policy-relevant effect size that can be communicated without specialized statistical background. Third, it surfaces scope conditions and limitations that are often overlooked: ecological inference issues, potential confounding by age structure, socioeconomic status, rurality/urban form, or food and activity environments, and the fact that estimates reflect associations at one point in time.

The paper is organized as follows: the Data section describes the dataset and how the analysis file was prepared; the Methods section outlines the regression model and reporting choices; the Results section presents the fitted model, the scatter plot with the regression line, and summary tables; the Discussion highlights interpretation and limitations; and the Reproducibility and References sections document the pipeline and sources.

Data

The units in this study are U.S. counties and county-equivalents (e.g., Louisiana parishes, Alaska boroughs). Each row corresponds to one county. The cleaned dataset used in the analysis contains 3,191 counties with complete information. Working at the county level provides a community-scale view of how behavior (physical inactivity) relates to health outcomes (obesity).

The data come from the 2023 County Health Rankings & Roadmaps (CHR&R) workbook (Harvard Dataverse). I use the Ranked Measure Data sheet, which compiles comparable county indicators derived largely from national surveys and small-area estimates. Although the workbook contains many indicators, this project focuses on two continuous measures expressed as percentages:

1. Adult Obesity (obesity_pct) — share of adults with BMI 30.
2. Physical Inactivity (inactivity_pct) — share of adults reporting no leisure-time physical activity.

To prepare the analysis file, I loaded the Ranked Measure Data sheet and retained standard identifiers (FIPS, state, county, year) together with the two target measures. Because the workbook includes many related fields (ranks, z-scores, confidence limits, numerators/denominators), I selected the value/percent columns only to keep scales consistent. Percentages encoded as text or proportions were parsed and, when necessary, rescaled so that both variables lie on a 0–100 scale. I then applied simple sanity checks—dropping rows with missing values or out-of-range percentages—and saved a clean analysis dataset in the project's data/ folder.

After cleaning, the modeling file contains 3,191 counties with complete information. Physical inactivity ranges from about 10% to 47%, and adult obesity ranges from about 18% to 53%, providing substantial between-county variation and a dense cloud of points suitable for linear regression and visualization. These indicators are derived from surveys and model-based estimation, so they carry sampling and measurement error that can vary with county size and response patterns. The analysis is ecological and cross-sectional; results describe county-level associations and should not be interpreted as individual-level causal effects.

Results

The fitted regression model shows a clear negative relationship between education and income at the neighborhood level. The estimated equation is:

$$\text{Obesity}_i = \beta_0 + \beta_1 \, \text{Inactivity}_i + \varepsilon_i,$$

where Obesity_i is the percent of adults with BMI ≥ 30 in county i; Inactivity_i is the percent of adults reporting no leisure-time physical activity in county i; β_0 is the intercept (predicted obesity when inactivity is zero); β_1 is the slope (change in obesity, in percentage points, per +1 pp in inactivity); and ε_i is the error term capturing unmeasured influences.

Estimation and reporting I fit the model by ordinary least squares (OLS) and report the estimated slope with a 95

Assumptions As with any SLR, the analysis relies on (1) **linearity** of the mean relationship, (2) **independence** of residuals across counties, (3) **approximately constant variance** of residuals (homoscedasticity; variance of ε_i roughly constant over Inactivity_i), and (4) **approximate normality** of residuals for interval estimates.

Scope and limitations. The model is **bivariate** and **ecological**: it summarizes an association at the county level and does not imply individual-level causation. Important covariates (e.g., age structure, income, rurality, built environment) are not included here but could be added in future extensions.

Software and workflow. All analyses were done in R (R Core Team 2024) using `readr` (Wickham, Hester, and team 2024), `dplyr` (Wickham, François, et al. 2024), `ggplot2` (Wickham, Chang, et al. 2024; Wickham 2016), and `broom` (Robinson, Hayes, and Couch 2024). A preprocessing script converts the CHR&R workbook into `data/county_health_model.csv`, and a modeling script fits the regression and generates the figure and tables used in the paper.

The fitted regression shows a clear positive association between county physical inactivity and adult obesity. The estimated equation is

$$\widehat{\text{Obesity}} \;=\; 17.81 \;+\; 0.714 \times \text{Inactivity}.$$

Here, the intercept of about **18** represents the predicted adult obesity percentage at zero inactivity (a hypothetical baseline). The slope indicates that for each **one–percentage-point** increase in inactivity, the predicted adult obesity rate increases by about **0.71** percentage points (**95% CI:** 0.70 to 0.73; $p < 0.001$). The model explains roughly **62.6**% of the between-county variation in obesity (**n = 3191**).

I test the null hypothesis (H_0: _1 = 0) against the two-sided alternative (H_1: _1 0). With the estimated slope, confidence interval, and p-value reported above, there is strong evidence of a positive association between inactivity and obesity at the county level.

Table 1: Regression results for Adult Obesity on Physical Inactivity.

term	estimate	std.error	statistic	p.value
(Intercept)	17.811	0.256	69.568	0
inactivity_pct	0.714	0.010	73.051	0

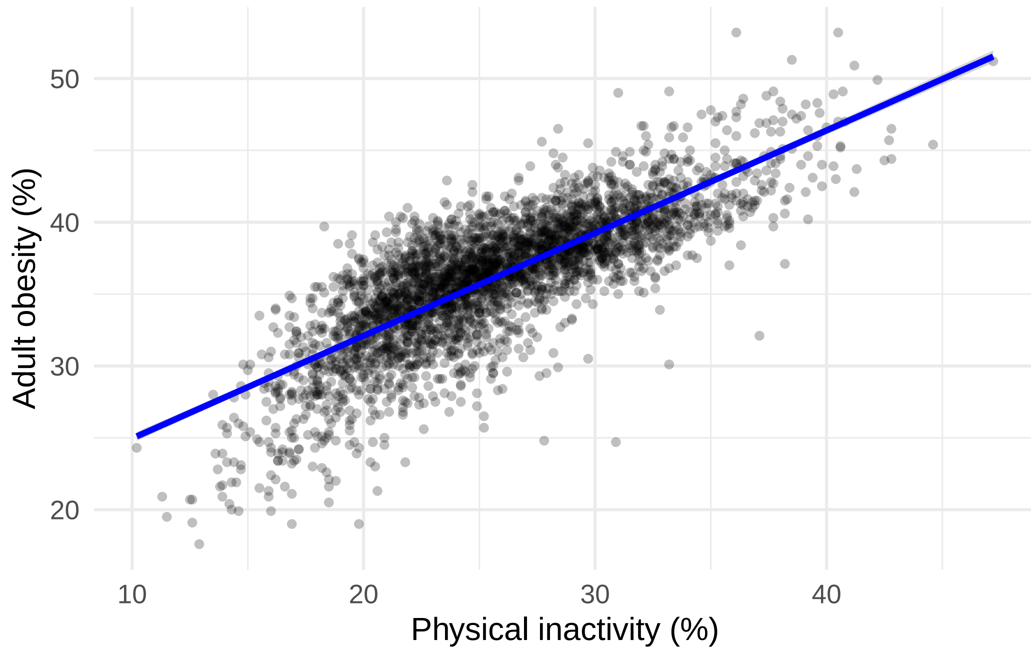

Figure 1: Adult Obesity (%) vs Physical Inactivity (%) — U.S. Counties (2023). Points are counties; the line is the OLS fit with a 95% confidence band.

References

County Health Rankings & Roadmaps. 2023. "County Health Rankings & Roadmaps: 2023 Ranked Measure Data." https://www.countyhealthrankings.org/.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2024. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://broom.tidymodels.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and PBC Posit Software. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://ggplot2.tidyverse.org/.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, Davis Vaughan, and PBC Posit Software. 2024. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org/.

Wickham, Hadley, Jim Hester, and posit team. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org/.