

## Summary of Hypotheses and Results

### Hypothesis 1: Bivariate Relationship

Hypothesis Statement: There is a statistically significant correlation between food desert prevalence and county-level obesity rates.

Statistical Test: Simple Linear Regression (Model 1)

Result: NOT REJECTED ✓

Evidence:

- Food desert prevalence shows a statistically significant positive correlation with obesity rates ( $r = +0.1417$ ,  $p < 0.001$ ). Model 1 regression results confirm this relationship:

Statistic	Value
Food Desert Coefficient ( $\beta$ )	+0.0268
Standard Error	0.003
t-statistic	8.007
p-value	<0.001
R-Squared ( $R^2$ )	0.020

- Interpretation: Each 1% increase in food desert prevalence is associated with a 0.027% increase in obesity rate. This relationship is highly statistically significant ( $p < 0.001$ ), indicating strong evidence against the null hypothesis of no relationship.
- Conclusion: Hypothesis 1 is SUPPORTED. A statistically significant bivariate relationship exists between food deserts and obesity. However, the low  $R^2$  value (0.020) indicates that food deserts alone explain only 2% of obesity variation, suggesting that other factors are critical.

### Hypothesis 2: Controlled Relationship with Socioeconomic Factors

Hypothesis Statement: Higher food desert prevalence predicts higher obesity rates after controlling for socioeconomic variables (income, poverty, education).

Statistical Test: Multiple Linear Regression (Model 2)

Result: REJECTED X

Evidence:

- When socioeconomic controls are added to the model, the food desert effect becomes statistically non-significant and changes direction:

Variable	Model 1	Model 2	Interpretation
Food Desert Coefficient	+0.0268***	-0.0029 (ns)	110% reduction; non-significant

- Model 2 Results - Controlling Variables:

Variable	Coefficient	p-value	Significance
Food Desert Prevalence	-0.0029	0.261	NOT SIGNIFICANT
Poverty Rate (%)	+0.1633	<0.001	***HIGHLY SIGNIFICANT
Education (Bachelor's+)	-0.2632	<0.001	***HIGHLY SIGNIFICANT
Median Income (\$)	-5.17e-6	0.416	NOT SIGNIFICANT

- The model fit improves dramatically with controls:  $R^2$  increases from 0.020 (Model 1) to 0.504 (Model 2), indicating that socioeconomic factors explain 50% of obesity variation.
- Critical Finding - Confounding Effect: The bivariate food desert-obesity relationship observed in Model 1 is SPURIOUS and driven by confounding. Counties with more food deserts tend to be poorer and less educated, and it is THESE socioeconomic factors—not food access itself—that drive obesity differences.
- Conclusion: Hypothesis 2 is NOT SUPPORTED in its original form. The food desert effect does NOT persist after controlling for socioeconomic factors ( $p = 0.261$ ). This is a critical finding that challenges the assumption that food deserts are an independent cause of obesity. Instead, food deserts are symptomatic of broader socioeconomic disadvantage.

### Hypothesis 3: Moderation by Socioeconomic Factors

Hypothesis Statement: Socioeconomic factors (particularly poverty) moderate the relationship between food desert prevalence and obesity rates. The strength of the food desert-obesity association varies depending on poverty levels.

Statistical Test: Moderated Regression with Interaction Terms (Model 3)

Result: NOT REJECTED ✓

Evidence:

- Model 3 includes an interaction term (Food Desert × Poverty) to test whether poverty modulates the food desert effect. Results are:

Variable	Coefficient	p-value	Interpretation
Food Desert (main effect)	-0.1325	0.042	Significant at low poverty levels
Food Desert × Poverty (interaction)	+0.2079	<0.001	***HIGHLY SIGNIFICANT

- The interaction coefficient is positive and highly significant ( $\beta = +0.2079$ ,  $p < 0.001$ ), indicating that poverty amplifies the food desert-obesity relationship.
- Practical Meaning: In LOW-POVERTY counties, food deserts have minimal or even negative association with obesity (better food access does not substantially improve outcomes if other health determinants are poor). In HIGH-POVERTY counties, food deserts show a stronger positive association with obesity, suggesting that in economically disadvantaged areas, geographic barriers to food access compound other health risks.
- Model Improvement:  $R^2$  improves from 0.504 (Model 2) to 0.507 (Model 3), a modest but statistically significant improvement that demonstrates the value of examining conditional effects.
- Conclusion: Hypothesis 3 is SUPPORTED. Poverty significantly moderates the food desert-obesity relationship. This is a novel finding that demonstrates the importance of context in understanding geographic health disparities. Food desert interventions will likely be most effective in economically disadvantaged communities.

Summary Table: All Hypotheses

Hypothesis	Test	Result	p-value / Evidence
H1: Bivariate relationship	Model 1	Supported ✓	$r = +0.14$ , $p < 0.001$
H2: Effect persists with controls	Model 2	Rejected X	$p = 0.261$ (non-sig)
H3: Poverty moderates effect	Model 3	Supported ✓	interaction $p < 0.001$

## Project Summary and Reflection

### Project Overview

This capstone project investigated the relationship between food desert prevalence and obesity rates across 3,129 U.S. counties, examining whether geographic barriers to food access explain differences in health outcomes. The analysis employed three regression models of increasing complexity to test this hypothesis while controlling for socioeconomic confounders and examining conditional effects.

### Key Findings

- Food deserts and obesity are correlated: A weak but statistically significant bivariate correlation exists ( $r = 0.14$ ,  $p < 0.001$ ), supporting the notion that food access and health outcomes are related.
- Socioeconomic factors dominate the relationship: When poverty, income, and education are controlled, the food desert effect disappears ( $p = 0.261$ ). This indicates that the bivariate relationship is spurious driven by confounding rather than a direct causal mechanism. Education ( $\beta = -0.263$ ) and poverty ( $\beta = +0.163$ ) are 10× more predictive of obesity than food deserts.
- Context matters: Moderation by poverty: Poverty significantly moderates the food desert-obesity relationship (interaction  $p < 0.001$ ). Food deserts have stronger effects in high-poverty counties, suggesting that geographic barriers compound disadvantage in economically fragile areas.
- Model comparison reveals confounding:  $R^2$  increases from 0.020 (Model 1) to 0.504 (Model 2) when socioeconomic controls are added, a 25-fold improvement that visually demonstrates the importance of statistical control in geographic health disparities research.

### Major Challenges Encountered

- Data Integration Across Geographic Scales: Food desert data was available at census tract level (11-digit FIPS codes) while health outcomes were at county level (5-digit FIPS codes). Aggregating census tract data to the county level required careful coding to preserve data integrity and ensure accurate prevalence calculations. The solution: extracting the first 5 digits of census tract IDs, grouping by county, and calculating the percentage of tracts meeting the "Low Income and Low Access" criterion.
- Confounding and Spurious Relationships: The bivariate analysis suggested food deserts strongly predict obesity, but this relationship was entirely explained by

- socioeconomic confounders. Recognizing and properly testing for confounding required moving beyond simple correlation to multivariate analysis. This highlights a critical lesson: geographic associations in health data often reflect underlying social inequities rather than direct causal mechanisms.
- Moderation Effects and Interpretation: While the moderation effect (poverty × food desert interaction) was statistically significant, its practical magnitude was modest ( $\Delta R^2 = 0.003$ ). Distinguishing between statistical and practical significance required careful interpretation and presentation of findings in context.
  - Large Sample Size and Multiple Comparisons: With N = 3,129 counties, even small effects achieve statistical significance. This required careful attention to effect sizes (standardized coefficients) rather than relying solely on p-values to assess the importance of findings.

### Lessons Learned

- Statistical Significance ≠ Practical Importance: Model 1 showed statistical significance ( $p < 0.001$ ) but explained only 2% of variance. This taught the importance of always examining both p-values and effect sizes and understanding that statistical tools can be misleading without interpretation grounded in domain knowledge.
- Confounding is Invisible in Bivariate Analysis: The dramatic difference between Model 1 and Model 2 illustrates how confounding can mask true relationships. Without properly testing multivariate models, researchers could wrongly attribute obesity to food access rather than socioeconomic disadvantage. This reinforced the importance of theory-driven variable selection and comprehensive control variable inclusion.
- Context Shapes Relationships: The moderation finding demonstrates that geographic and social relationships are heterogeneous. Interventions should not be one-size-fits-all; food desert programs may be most effective in high-poverty areas where barriers compound. Future work should examine regional variation, rural vs. urban dynamics, and other sources of heterogeneity.
- Data Quality and Integration are Critical: Careful attention to column names, FIPS code formatting, and geographic unit alignment was essential. Small errors in data preparation would have propagated through all subsequent analyses. This reinforced best practices in data cleaning, validation, and documentation.
- Reproducibility Requires Transparent Code: Python-based analysis with version control and inline documentation ensured that all steps could be audited and

reproduced. This approach is far superior to manual Excel manipulations and is essential for research integrity.

## **Project Contribution and Future Direction**

This project contributes to the field in several important ways:

- It provides empirical evidence that food desert-obesity relationships documented in prior research are largely confounded by socioeconomic factors, suggesting that policy responses must address root causes (poverty, education) rather than symptoms (food access alone).
- The moderation analysis identifies an important conditional relationship: food deserts matter more in high-poverty areas. This suggests a nuanced policy approach where food desert interventions are most effective in economically disadvantaged communities.
- Methodologically, it demonstrates best practices in multivariate analysis, confounding assessment, and interaction testing—skills applicable to many research questions beyond this specific problem.

Future work building on this analysis could include:

- Mechanisms: Examine potential pathways. Do food deserts affect diet quality, food costs, dietary diversity, or time spent on food acquisition? Which pathways matter most?
- Regional variation: Are food desert effects similar in rural vs. urban areas? In different U.S. regions? Are regional policy interventions more effective?
- Additional variables: Healthcare access, SNAP participation rates, vehicle ownership, built environment factors (walkability, gym access) could further refine our understanding.
- Heterogeneous effects: Use machine learning or causal forests to identify which subpopulations benefit most from food desert interventions.
- Qualitative research: Interviews and focus groups in high-poverty, high-food-desert areas could illuminate the lived experience and identify barriers beyond data captured in county-level statistics.

## Real-World Implications

This analysis addresses critical public health and social policy questions with implications for healthcare systems, urban planning, education policy, economic development, and health equity. The following sections examine how these findings inform responses to real-world problems affecting millions of Americans.

### The Public Health Crisis: Obesity and Its Burden

Obesity is one of the most pressing public health challenges in the United States. According to CDC data, approximately 42% of American adults are obese, with significant disparities across race, ethnicity, and socioeconomic status. This analysis found that obesity rates in the county sample ranged from 17.6% to 52.2%, representing a three-fold variation across the country. This geographic variation demands explanation and targeted intervention.

Obesity contributes to an estimated 300,000 excess deaths annually in the U.S. and costs the healthcare system over \$147 billion per year in direct medical expenses, plus an additional \$66 billion in indirect costs (lost productivity, disability). At the individual level, obesity increases risk for:

- Type 2 diabetes (risk increased 80-fold in obese individuals)
- Cardiovascular disease (heart attack, stroke)
- Hypertension (high blood pressure)
- Certain cancers (endometrial, breast, colon)
- Sleep apnea and respiratory problems
- Joint problems, arthritis, and mobility limitations
- Mental health challenges including depression and anxiety

The health burden of obesity is not equally distributed. The highest obesity rates are concentrated in rural areas, the South, and communities with high poverty rates. In other words, the areas where this analysis found food deserts to be prevalent. However, my analysis reveals an important finding: the problem is not food deserts alone, but the broader socioeconomic context.

## Economic Development and Living Wages

This analysis found that poverty rate is a strong predictor of obesity (coefficient: +0.1633, p < 0.001). Economic insecurity shapes health outcomes through multiple pathways:

- Food Affordability: In low-income households, budget-friendly foods tend to be calorie-dense, highly processed options. Fresh produce, lean proteins, and whole grains are more expensive per calorie, making poverty a barrier to healthy eating.
- Time Poverty: Low-wage jobs often involve long hours, multiple jobs, and inflexible schedules. This "time poverty" limits time for meal preparation, exercise, sleep, and self-care—all health-protective behaviors.
- Stress and Health: Economic insecurity is a chronic stressor linked to elevated cortisol, inflammatory markers, and increased weight gain. The psychological burden of poverty affects health directly through physiological pathways.
- Neighborhood Quality: Poverty concentrates in neighborhoods with fewer resources, higher crime, less safe recreation spaces, and limited access to services. These environmental factors limit physical activity and recreational opportunities.

These mechanisms suggest several policy approaches to address poverty's health effects:

- Living Wage Standards: Cities and states should establish living wage requirements that provide economic security to workers. Research shows that inadequate income is a primary driver of stress and poor health outcomes.
- Job Quality and Benefits: Beyond wages, job quality matters. Access to health insurance, paid sick leave, flexible scheduling, and retirement benefits all contribute to health and economic security.
- Social Safety Nets: SNAP (food assistance), TANF (cash assistance), housing assistance, and other programs are essential. However, eligibility barriers and benefit adequacy must be addressed.
- Economic Opportunity: Job training programs, business development services, and capital access for small enterprises in low-income communities create pathways to self-sufficiency and wealth building.

- Childcare and Early Education: Access to affordable, quality childcare removes barriers to employment for parents, particularly mothers, and provides developmental benefits for children.