Food Supply and Death Attributed to Obesity

Ken Ye

2023-12-17

Introduction

With an alarming 41.9% of the U.S. population classified as obese, as revealed by the 2021 National Health and Nutrition Examination Survey (NHANES) [1], the implications of this health crisis become undeniably significant. However, the impact becomes even more pronounced when considered globally, where obesity contributed to 8% of deaths in 2017 — an alarming surge from 4.5% in 1990 [2].

Obesity, beyond its visual manifestations, is intricately linked to severe health conditions, including heart disease, stroke, type 2 diabetes, and various cancers. These conditions stand as leading causes of preventable, premature deaths. The economic burden is equally staggering, with the estimated annual medical cost of obesity in the United States reaching nearly $173 billion in 2019 dollars. Notably, medical costs for adults grappling with obesity were $1,861 higher than those for individuals maintaining a healthy weight [1].

As we delve into the complex web of factors contributing to obesity, the role of food emerges as a central player. Continuous overconsumption, a potential precursor to weight gain and obesity, emphasizes the critical importance of understanding food composition. Certain foods, due to their nutrient content, can significantly influence overall health outcomes.

However, the narrative extends beyond food alone. Economic status emerges as a pivotal consideration, acting as a covariate in my analysis. Death rates attributed to obesity exhibit variations among countries, and a fundamental differentiator lies in their economic standing. Developed countries, buoyed by high incomes, may boast advanced medical systems capable of mitigating obesity-related deaths. Paradoxically, these countries may also grapple with higher obesity prevalence. On the contrary, developing nations, constrained by lower incomes, may face a different set of challenges, potentially emphasizing malnutrition over obesity.

In this analysis, the focal point is discerning the effects of a country's food supply and economic status on death rates attributed to obesity. The implications of this study are far-reaching, offering valuable insights for policymakers eager to curb obesity-related deaths and aiding consumers in crafting diets resilient to the detrimental effects of obesity.

A dual-stage methodology will be employed. Initially, simple linear regression sheds light on the potential impact of individual food groups and a country's economic status. This analysis guides us towards identifying statistically significant associations. Subsequently, the methodology advances to Lasso regression, a more sophisticated model aimed at pinpointing the most influential variables while mitigating the risk of overfitting.

Data

Data Sources

Three separate datasets, totaling 26 predictor variables and 1 response variable, are used in this analysis:

1) Daily caloric supply (OWID based on UN FAO & historical sources)

Predictor variables (after renaming): `Miscellaneous`, `Alcohol`, `Animal.fat`, `Vegetable.oils`, `Oilcrops`, `Fish.and.seafood`, `Sugar.crops`, `Sugar.sweeteners`, `Starchy.roots`, `Meat.other`, `Meat.sheep.and.goat`, `Meat.pig`, `Meat.poultry`, `Meat.beef`, `Eggs`, `Milk`, `Nuts`, `Fruit`, `Vegetables`, `Pulses`, `Cereals.other`, `Barley`, `Maize`, `Rice`, `Wheat` — each one (25 total) is a food group available for consumption, continuous, unit (kcal per day per capita)

Source: Our World in Data based on UN FAO & historical sources — processed by Our World in Data

Link: https://www.fao.org/faostat/en/#data/FBS

2) World Bank income classification

Predictor variable (after renaming): `Income.group` — nominal, four categories (Low-income countries, Lower-middle-income countries, Upper-middle-income countries, High-income countries)

Source: World Bank (2022) — processed by Our World in Data

Link: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

3) Share of total deaths that are from all causes attributed to high body-mass index, in both sexes aged age-standardized

Response variable (after renaming): `Death.obesity` — continuous, unit (%)

Source: IHME, Global Burden of Disease Study (2019) — processed by Our World in Data

Link: https://ourworldindata.org/obesity

Data Cleaning & Merging

To ensure chronological congruence in the analysis, I selected the latest year (2019) for which data was available in all three datasets and excluded any observations for entities that are not countries (e.g. Africa), focusing on individual countries. Renaming variables enhanced legibility. Finally, I merged the three datasets using the inner join method on the variable `Country`. The resulting merged dataset, named `df`, comprises 181 observations and 28 columns.

Missing Data Imputation

During the examination of the data, I encountered missing values in certain food groups across countries. Removing these entries would significantly reduce the dataset, leaving only 79 out of 181 countries (with 102 observations deleted due to missingness) with 50 degrees of freedom left for model fitting if all columns are utilized. This approach would not be ideal for accurately representing global patterns or obtaining robust model estimates.

To address this challenge, I opted for imputation using the "predictive mean matching"(PMM) method from the `mice` package. PMM ensures that imputed values are drawn from observed values with similar predicted values, preserving the distributional characteristics of the data. This method allows us to utilize the full dataset, providing a more comprehensive and representative view of the relationships between food supply patterns and obesity death rates across diverse countries. Imputing missing values with PMM contributes to the validity of our inferences and enhances the generalizability of my findings to a broader global context.

EDA

Obesity-attributed Death Visualization

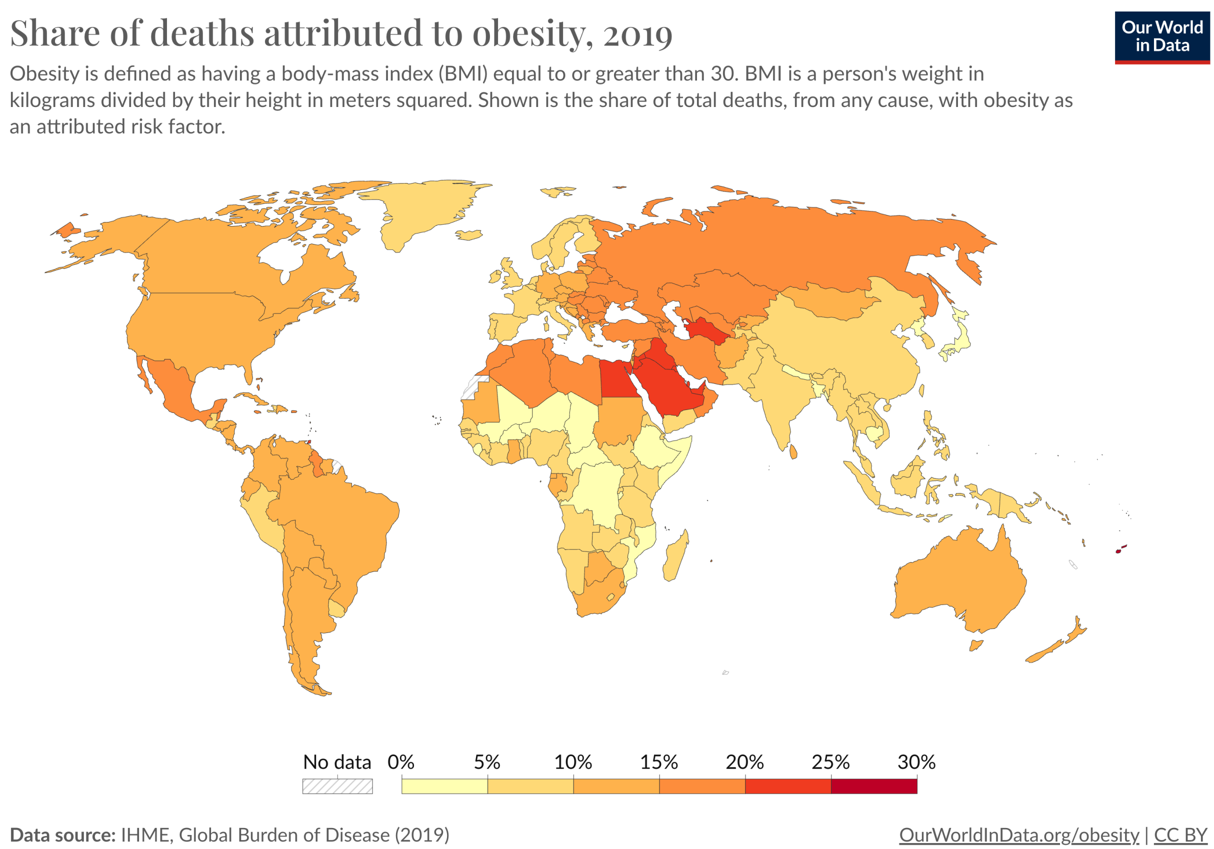


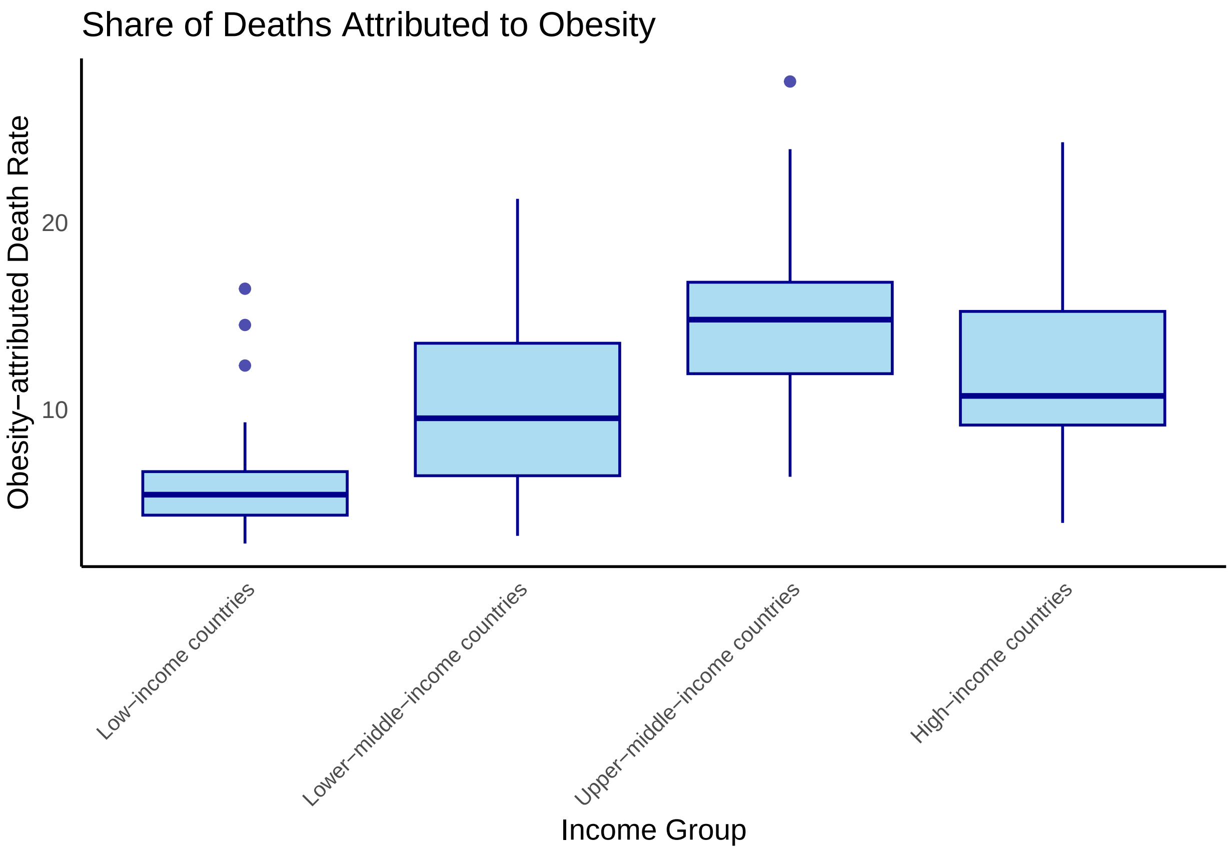
Figure 1: Share of Deaths Attributed to Obesity Across the Globe

The distribution of obesity-attributed death varies significantly on a global scale, as illustrated in the above map.

In numerous middle-income nations, notably in Eastern Europe, Central Asia, North Africa, and Latin America, over 15% of deaths were linked to obesity. This pattern likely arises from a high prevalence of obesity coupled with comparatively poorer overall health and healthcare systems, especially when compared to high-income countries experiencing similar levels of obesity.

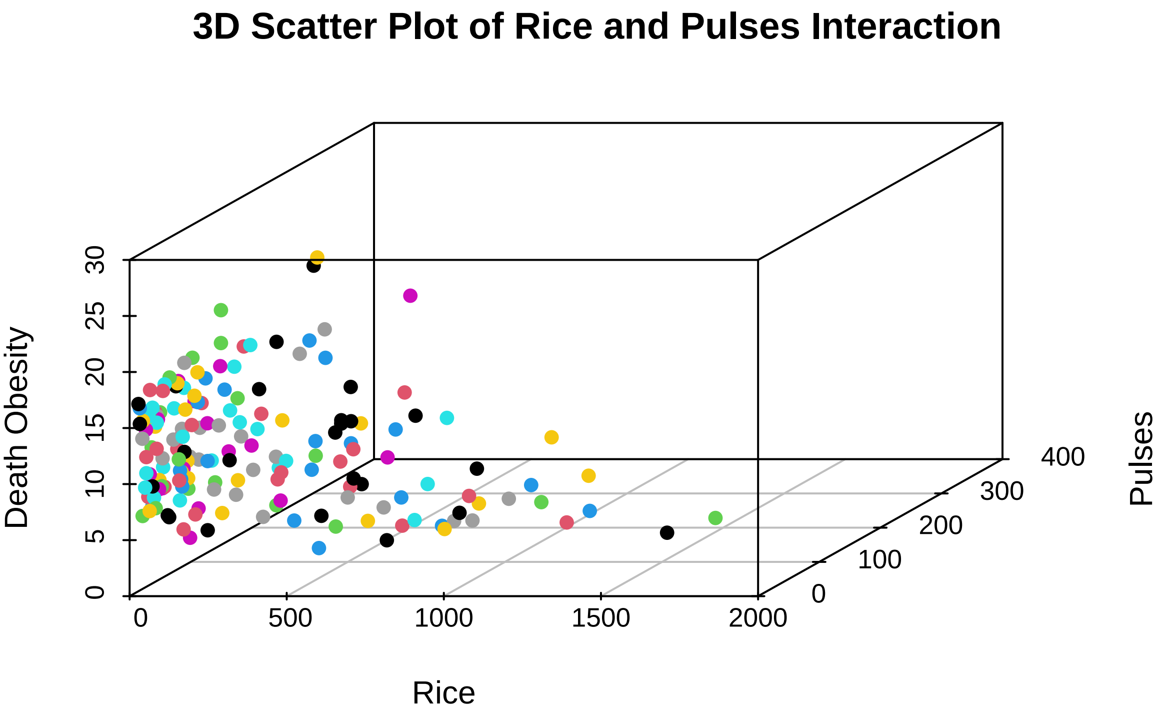
Contrastingly, in most high-income countries, this proportion ranges between 8% to 10%, approximately half of that observed in many middle-income countries. Notable exceptions among affluent nations are Japan and South Korea, where only around 5% of premature deaths are attributed to obesity.

Turning to low-income countries, particularly those in Sub-Saharan Africa, obesity accounts for less than 5% of deaths, underscoring a distinct epidemiological profile in these regions.



The graph above, utilizing 2019 data, provides a clearer depiction of how the obesity-attributed death rate varies across countries' economic statuses. Notably, as a country progresses through income group classifications from Low-income to Lower-middle-income to Upper-middle-income, there is an observable increase in the obesity-attributed death rate. Intriguingly, when a country transitions from Upper-middle-income to High-income, the obesity-attributed death rate begins to decrease once more. This pattern adds an intriguing layer to the analysis, prompting further exploration into the nuanced relationship between economic status and obesity-related mortality.

Exploration of Interaction Effects



During meal consumption, certain food pairings emerge as common occurrences — rice (`Rice`) paired with beans (`Pulses`), chicken (`Meat.poultry`) alongside broccoli (`Vegetables`), and steak (`Meat.beef)` accompanied by mashed potatoes (`Starchy.roots`), to name a few. Exploring potential interaction effects between the supply of these food groups could shed light on nuanced dietary patterns and their impact on obesity-attributed death. However, an examination of the coefficients for these interaction terms reveals a lack of statistical significance. Consequently, the exploration of interaction effects is not pursued further in the subsequent models.

Methodology

The 181 observations are randomly divided into 80% training and 20% testing sets.

Simple Linear Regression Model

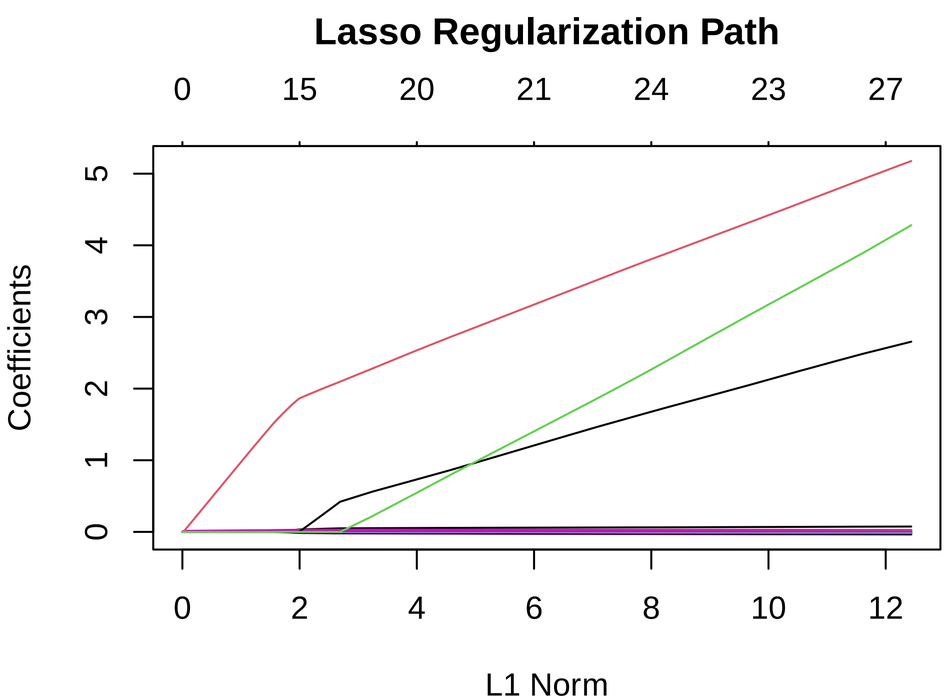
The simple linear regression model is employed to predict the obesity-attributed death rate based on the supply of 25 food groups and the country's income group. The model exhibited a Multiple R-squared of 0.6466, indicating a reasonable level of explained variability. Significant predictors included specific food groups such as fish and seafood, sheep and goat, pig, poultry, beef, and wheat. Among income groups, lower-middle-income countries, upper-middle-income countries, and high-income countries also demonstrated notable effects. The F-statistic was significant (p-value: 2.181e-15), affirming the model's overall significance. In addition, while checking for multicollinearity issues, all variables except for `Upper-middle-income countries` (VIF = 5.4) and `High-income countries` (VIF = 8.9) had VIF values < 5, indicating low or no multicollinearity issues. This initial analysis provides valuable insights into factors influencing obesity-related mortality, paving the way for further refinement and exploration through techniques like Lasso regression.

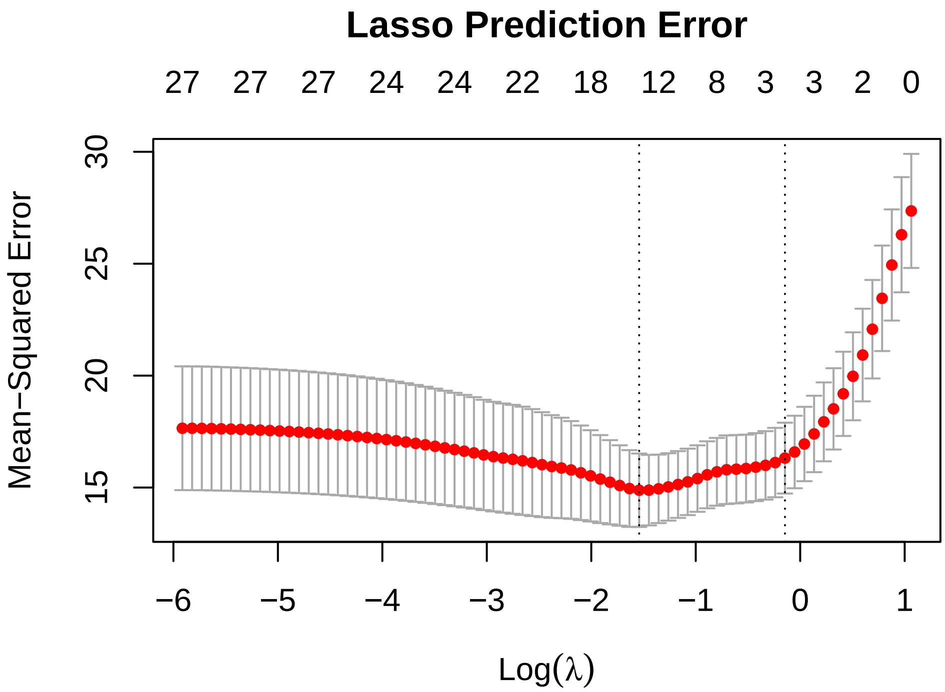
Lasso Regression Model

Lasso regression provides several advantages for refining predictive models. One key feature is the introduction of a regularization term that encourages sparsity in coefficient estimates, facilitating automatic variable selection. This proves particularly beneficial when dealing with a substantial number of predictors, such as the 26 predictors in this case, where prior knowledge about their significance may be limited. Lasso tends to shrink certain coefficients to zero, effectively excluding irrelevant variables and enhancing model interpretability.

In addition to variable selection, Lasso addresses multicollinearity issues by favoring a unique subset of correlated predictors. This not only contributes to model stability but also ensures interpretability by focusing on the most relevant information. The regularization incorporated in Lasso goes beyond enhancing model generalization; it also promotes robustness, making the model more resilient to fluctuations in the data and, consequently, improving predictive accuracy.

Furthermore, the importance of standardization cannot be overstated, especially for predictors related to food groups. Standardization ensures that all variables are on the same scale, crucial when dealing with diverse food supplies in varying absolute quantities. Since some food groups may have low consumption rates, standardization aids in providing a fair comparison between different predictors, contributing to the overall effectiveness of the Lasso regression model.





To fit the Lasso model, a ten-fold cross-validation (CV) approach is employed to identify the optimal lambda value, which is determined to be 0.23. This lambda value is subsequently utilized to construct the final model using the entire dataset.

Model Comparison

When comparing the prediction accuracy of the Lasso model with the simple linear regression model, a substantial improvement is evident. The testing Mean Squared Error (MSE) decreased significantly, dropping from 44.6 in the simple linear regression model to 7.8 in the Lasso model. This noteworthy reduction underscores the enhanced performance of the Lasso model, emphasizing its superior predictive capabilities in capturing the underlying patterns within the data.

Results

Final Model (Lasso)

	Coefficient
(Intercept)	5.2330
Oilcrops	0.0089
Fish.and.seafood	-0.0192
Sugar.crops	0.0403
Sugar.sweeteners	0.0032
Meat.sheep.and.goat	0.0162
Meat.pig	-0.0036
Meat.poultry	0.0261
Meat.beef	-0.0161
Fruit	0.0006
Pulses	-0.0026
Cereals.other	-0.0051
Wheat	0.0065
Income.groupLower-middle-income countries	0.1101
Income.groupUpper-middle-income countries	1.9166

Model Interpretation

The table above presents the variables selected by the Lasso model along with their corresponding coefficients. These coefficients indicate the strength and direction of the relationship between each predictor and the obesity-attributed death rate. Here is an interpretation of the selected variables:

Intercept: The intercept of 5.48% represents the predicted obesity-attributed death rate when the supply of all food groups is theoretically zero, and the country is classified as part of the High-income group. It's essential to note that this scenario is not practically meaningful, as complete absence of food supply is unrealistic. The intercept serves as a baseline reference point for the model but should be interpreted with caution in the context of food-related variables.

Oilcrops: The positive coefficient for Oilcrops indicates that a one-unit increase in the supply of oil-rich crops, measured in kilocalories per day per capita, is associated with an estimated 0.80% increase in the obesity-attributed death rate, holding other variables constant. This relationship suggests that a higher intake of oil-rich crops, such as certain seeds and nuts, may contribute to a diet with a higher fat content, potentially influencing obesity-related health issues.

Fish and Seafood: The negative coefficient for Fish and Seafood suggests that a one-unit increase in the supply of fish and seafood is associated with an estimated 1.72% decrease in the obesity-attributed death rate, holding other variables constant. This aligns with the established health benefits of fish consumption, as fish is a rich source of omega-3 fatty acids, which may contribute to improved cardiovascular health and a potential reduction in obesity-related risks.

Sugar Sweeteners: The positive coefficient for Sugar Sweeteners indicates that a one-unit increase in the supply of sugar sweeteners is associated with an estimated 0.35% increase in the obesity-attributed death rate, holding other variables constant. This suggests that higher supplies of sugar sweeteners, often found in processed foods and beverages, may contribute to an increased risk of obesity-related health issues due to excessive sugar consumption.

Meat Sheep and Goat: The positive coefficient for Meat Sheep and Goat suggests that a one-unit increase in the supply of sheep and goat meat is associated with an estimated 0.0145% increase in the obesity-attributed death rate, holding other variables constant. This relationship may be explained by the higher saturated fat content in these meats, which, when consumed excessively, may contribute to adverse health outcomes related to obesity.

Meat Pig: The negative coefficient for Meat Pig indicates that a one-unit increase in the supply of pig meat is associated with an estimated 0.0032% decrease in the obesity-attributed death rate, holding other variables constant. This may be due to a potential protective effect, as pig meat is often leaner compared to some other meats, and moderate consumption of lean protein may be associated with better health outcomes.

Meat Poultry: The positive coefficient for Meat Poultry suggests that a one-unit increase in the supply of poultry meat is associated with an estimated 0.0242% increase in the obesity-attributed death rate, holding other variables constant. This relationship may be associated with the consumption of processed poultry products, such as deli meats and deep-fried products, which can be high in unhealthy fats and additives, potentially contributing to obesity-related health issues.

Meat Beef: The negative coefficient for Meat Beef indicates that a one-unit increase in the supply of beef is associated with an estimated 0.0154% decrease in the obesity-attributed death rate, holding other variables constant. This suggests a potential protective effect, possibly because lean beef can be part of a balanced diet, providing essential nutrients without excessive unhealthy fats.

Fruit: The positive coefficient for Fruit indicates that a one-unit increase in the supply of fruits is associated with an estimated 0.0004% increase in the obesity-attributed death rate, holding other variables constant. This relationship may be explained by the fact that while fruits contain essential vitamins and fiber, excessive consumption of certain fruits high in natural sugars may contribute to caloric intake and, in turn, obesity-related risks.

Pulses: The negative coefficient for Pulses suggests that a one-unit increase in the supply of pulses is associated with an estimated 0.0013% decrease in the obesity-attributed death rate, holding other variables constant. This potential protective effect may be attributed to the high fiber and plant protein content in pulses, contributing to satiety and a balanced diet, reducing the risk of obesity-related health issues.

Cereals Other: The negative coefficient for Cereals Other indicates that a one-unit increase in the supply of other cereals is associated with an estimated 0.0021% decrease in the obesity-attributed death rate, holding other variables constant. This relationship may be associated with the inclusion of whole grains in this category, known for their health benefits, including promoting satiety and contributing to better weight management.

Wheat: The positive coefficient for Wheat indicates that a one-unit increase in the supply of wheat is associated with an estimated 0.0063% increase in the obesity-attributed death rate, holding other variables constant. This suggests a potential increase in obesity-related health risks with higher wheat supply, possibly due to the consumption of refined wheat products lacking the nutritional benefits of whole grains.

Income Groups:

Lower-middle-income countries: The positive coefficient for Lower-middle-income countries indicates that being classified as part of this income group is associated with an estimated 0.0289% increase in the obesity-attributed death rate, holding other variables constant. This relationship may be influenced by various factors, including changes in dietary patterns, lifestyle, and healthcare infrastructure that contribute to obesity-related health risks.

Upper-middle-income countries: The significant positive coefficient for Upper-middle-income countries suggests a more substantial impact, with an estimated 1.8862% increase in the obesity-attributed death rate, holding other variables constant. This may reflect a complex interplay of dietary, lifestyle, and socioeconomic factors in upper-middle-income countries contributing to heightened obesity-related health risks.

Implications

The findings from this analysis hold substantial implications for policymakers and consumers, providing actionable insights to combat the escalating issue of obesity-related deaths. For policymakers, the identified significant predictors offer a roadmap for targeted intervention strategies. The positive coefficient for

`Lower-middle-income countries`, for instance, signals a need for tailored public health initiatives in these economic contexts. Policymakers can leverage this information to design impactful public health campaigns, promoting the consumption of healthier food choices, such as Fish and Seafood, to mitigate the adverse effects of obesity. The insights into the economic considerations, as highlighted by the Income Group coefficients, allow policymakers to allocate resources efficiently, considering economic disparities in addressing obesity-related health risks. Additionally, consumers can benefit from this analysis by making informed dietary choices. Understanding the impact of specific food groups, such as Sugar Sweeteners, encourages moderation in the consumption of processed foods, while the negative coefficient for Fish and Seafood suggests substituting red meat with fish for potential health benefits. Moreover, the nuanced effects of different meat types provide consumers with insights for achieving a balanced diet. By incorporating whole grains and fiber-rich foods, consumers can contribute to better weight management. The awareness of economic disparities influencing obesity-related death rates prompts consumers to consider the broader socioeconomic context and advocate for systemic changes. In summary, these findings empower both policymakers and consumers with evidence-based information to address obesity-related deaths comprehensively.

Conclusion

This project explored the global impact of obesity-related deaths, focusing on the interplay of food supply, economic status, and mortality rates. Key findings highlighted the severity of obesity-related health crises and the economic burden. The analysis identified significant factors using a dual-stage methodology, revealing nuanced associations.

Economic status, represented by income groups, showed distinct links. Lower-middle-income countries were associated with increased death rates, emphasizing the need for targeted interventions. Upper-middle-income countries exhibited a significant rise in mortality rates, warranting further investigation.

Food groups played a crucial role, with positive associations found for oil-rich crops and sugar sweeteners, indicating potential health risks. Negative associations with fish and seafood suggested protective effects, guiding healthier dietary choices.

The Lasso regression model refined variable selection, emphasizing the impact of specific food groups and economic factors. These insights hold practical implications for policymakers in crafting targeted interventions and resource allocation. For consumers, the findings empower informed dietary choices to manage weight and enhance overall health.

In conclusion, this project contributes evidence-based insights to address the complex challenge of obesity-related deaths globally, providing a foundation for comprehensive strategies.

Limitations and Future Research

This analysis has certain limitations that should be acknowledged. Firstly, the use of food supply data may not precisely reflect consumption, especially in higher-income countries where food waste is prevalent. Access to an actual food consumption dataset would provide a more accurate basis for interpreting health outcomes, including obesity-related deaths.

Additionally, the study focused on food supply and income status as primary factors influencing a country's obesity-related death rate. However, the multifaceted nature of this outcome involves various factors such as physical activity and healthcare systems. Future research should expand the scope by incorporating additional variables to provide a more comprehensive understanding of the complex determinants of obesity-related deaths.

Citations

[1] https://www.cdc.gov/obesity

[2] https://ourworldindata.org/obesity