HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 4 (2023)

Analysis of Obesity Based on Various Indicators by Countries via Linear Regression

Athena Rhee[1] and Priyanjali Datta[#]

[1]Phillips Academy
[#]Advisor

ABSTRACT

Obesity, a prevalent disease, begins with the Third Industrial Revolution. Obesity not only lowers numerous individuals' quality of life, but also affects budgets in the health sector across the country and around the world. Obesity was once considered an exclusive problem among the rich: In general, high-quality meat is expensive, and plants (grains) have been relatively affordable. This proposition held true until the late 20th century, but today, obesity is no longer just a concern for the affluent. Factors such as climate change, supply-chain problems, high consumption demands, and more expensive materials such as feed and biofuels result in a price surge. Increased food prices have a direct impact on households, making it more difficult for the lower-income class to access nutritious foods. As a result, the poor reduce their consumption of nutritious foods such as fruits, vegetables, and grains and replace meals with inexpensive, artificial meat, and ultra-processed foods. This paper proposes policies by analyzing and predicting the degree of obesity using linear regression, a data mining technique, according to the development or welfare level of the country using various indicators by country. This paper refines and evaluates data through a preprocessing method after data collection and concludes by explaining the results.

1. Introduction

1.1 Motivation

According to the Global Burden of Disease, in 2017, overweight and obesity has become an epidemic issue to the point where more than 4 million people are affected, contributing to an upsurge in the global mortality rate. Each year, obesity rates have been increasing constantly. In 2019, about 38.2 million children under five were overweight or obese (World Health Organization [WHO], 2021). Data from 1975 to 2016 reveals that the prevalence of overweight or obese children and adolescents aged 5–19 increased more than four-fold from 4% to 18% worldwide (WHO, n.d.). Similarly, in 2016, more than 1.9 billion adults aged 18 years and older were overweight. Of these, more than 650 million adults were obese (WHO, 2021). Kelly et al. estimated that the global prevalence of overweight and obesity may surpass 57% by 2030. They reported that the estimated total numbers of overweight and obese adults in 2005 were 937 million and 396 million, respectively. These numbers were projected to become 1.35 billion and 573 million without adjusting for secular trends by 2030 (Kelly et al., 2008).

Overweight and obesity, formerly thought to be a problem exclusively in high-income nations, are now sharply increasing in low- and middle-income nations. The majority of overweight or obese children are found in developing countries, where obesity rates have outpaced developed nations by more than 30% (WHO, n.d). According to a study conducted in China involving 12,543 individuals over a 22-year period, age-adjusted obesity prevalence increased from 2.15% to 13.99% in males and females, ranging from 2.78 to 13.22% for women and from 1.46 to 14.99% for men, respectively (Chen et al., 2019; Jia, 2015). The overweight rate of African children under five years old has increased by 24% since 2000. As of 2019, almost half of Asian children under five years old were obese or

overweight. Since 2000, there has been a 24% increase in the percentage of African children under five years who are overweight. Nearly half of Asian children under the age of five were overweight or obese as of 2019 (Wariri et al., 2021).

Even though obesity is preventable, it has become a challenge globally. As the trends are rising continuously, controlling this health epidemic has become crucial. On an individualistic level, it is said that obesity is caused by the imbalanced consumption of energy-dense foods like fast foods, including loaded burgers, sodas, etc., or not having a good physical activity regime. However, if we move beyond these causes, we see the big picture of nationwide systemic causes. For instance, historic residential segregation due to racism has caused a lot of trouble for African American in the U.S, who are severely affected by obesity, which has led to several serious issues, including lack of access to health care, unequal employment opportunities, access to quality education, difficulty accessing healthy foods which leads to food insecurity, etc. Groups in other developing countries might be facing different types of systemic issues where unhealthy food consumption is not regulated, or whole foods are too expensive to purchase, so people are ready to consume inexpensive foods loaded with higher amounts of saturated fats, sugars, and empty calories. These problems are on the rise and, thus, need more targeted solutions. Moreover, recent media has called attention to the ways in which economic globalization, in particular trade liberalization, has aided in the spread of obesogenic goods to low and middle-income countries (LMICs), such as sugar-sweetened beverages and processed foods. People perceive Western diets as "modern food" and wish to consume such foods, moving away from their historical and cultural foods, which are almost always healthier (Fox et al., 2019).

1.2 Obesity

Obesity is defined and categorized using body mass index (BMI). Different BMI thresholds and assessments of body fat quantity and its distribution are utilized to indicate overweight and obesity in descriptive statistics. In the West, a value of 27 kg/m^2 was marked as the value above and beyond, which obesity was considered; this was later changed to 30 kg/m^2.

Regardless of the changes, these values were unsuitable for the Asian population due to their specific health implications and body compositions (Tan, 2004). In China, 28 kg/m^2 is used as the cut-off value for obesity in the Chinese population (Hou et al., 2008). For Indians, a BMI value of ≥25 kg/m^2 is used as the cut-off value for obesity (Behl & Misra, 2017).

Other experts argue that obesity must be defined as excessive fat accumulation in the body. Over the centuries, the human body has made specific changes to adapt to the conditions and store the surplus energy in the form of fat, to be used as stored energy in times of energy shortage (Leibel et al., 1995). To elaborate on the previous point, we can look at the example of edema. The belief that edema is caused by drinking too much water is misleading. In fact, edema is caused by disturbances in the water balance, leading to abnormal water accumulation in body cells. Likewise, several bodily functions work together and cause the uptake or expenditure of energy. The distribution of any single or more such bodily mechanisms can cause excess energy storage, eventually leading to obesity (Dhurandhar et al., 2021; Dhurandhar, 2022).

Obesity has become a crucial public health concern, even among infants. Obese children are at a greater risk of developing obesity-related diseases in the future. Previous evidence from research proves that maternal obesity affects a child's health adversely in the short and long term. Maternal factors affect childhood diseases, persisting into adulthood. During pregnancy and lactation, maternal obesity and weight gain lead to changes in microbiota composition and activity, influencing diversity in gut and breast milk. These alterations in the microbiota could be passed onto the offspring during delivery and lactation, affecting the infant's microbial colonization and the development of the immune system (Garcia-Mantrana & Collado,

2016). A beneficial strategy to reduce the risk of obesity and overweight in infants involves promoting a positive nutritional and microbial environment during the perinatal period. This can be achieved by targeting early-life microbiota modulation. Understanding and harnessing the interplay between maternal obesity, microbiota, and

infant health could lead to improvement in obesity prevention efforts and better health outcomes for future generations (Yao et al., 2020; Yao et al., 2021).

Contemporary health guidelines to control obesity are grounded in the fundamental physiological principle that fat buildup results from an energy imbalance between ingested and burned calories. To address this, one must rectify the inconsistency by either consuming fewer or burning more calories than what is taken in. The epidemic of obesity is made more vigorous by the ingestion of highly available energy-dense and low-nutrition food. A person consumes food based on the variables such as availability, cost, environment, etc. Due to these varying factors, it is problematic for many individuals to find the right balance (Yoo, 2018). According to a study conducted in 2017 to evaluate the coronary artery risk development in young adults, results showed that individuals who would consume more fast-food would weigh more (approximately 6 kg) than the individuals with significantly low fat-food intake (Duffy et al., 2017).

The complexity of these issues is compounded in individuals who exhibit a genetic susceptibility to fat accumulation. Factors such as the accumulation of lipid metabolites, inflammatory signaling, or other mechanisms impairing hypothalamic neurons could also contribute to obesity, potentially explaining the biological defense mechanism of elevated body fat mass. Moreover, these challenges are further exacerbated by societal factors such as sedentary lifestyles, unhealthy eating habits, and easy access to high-calorie, processed foods. Additionally, psychological factors such as stress, emotional eating, and lack of proper coping mechanisms can play a role in the development of obesity. Addressing this multifaceted issue requires a comprehensive approach, including promoting healthier lifestyles, early detection of genetic predispositions, and targeted interventions tailored to individual needs (Obri & Claret, 2019; Lin & Li, 2021).

1.3 Diseases Triggered by Obesity

Obesity increases the likelihood of various diseases and conditions which are linked to increased mortality. Obesity is linked with elevated rates of hypertension, stroke, and cardiovascular diseases. The development of these conditions in obese individuals encompasses various mechanisms. Specifically, pro-inflammatory and pro-thrombotic adipokines are believed to contribute to the heightened risk of cardiovascular disease. Additionally, obesity may lead to increased vascular volume, heightened arterial resistance, and the release of angiotensinogen from enlarged fat cells, thereby potentially contributing to elevated blood pressure. These cumulative factors highlight the critical importance of addressing obesity as a significant risk factor for a spectrum of cardiovascular issues (Yatsuya et al., 2010; Schnabel et al., 2013).

The global increase in obesity has contributed to an increased prevalence of type 2 diabetes (Klein et al., 2022). Obesity emerges as a prominent predictor of type 2 diabetes. The onset of type 2 diabetes centers around alterations in either or both key factors: insulin sensitivity and insulin secretion. As per the evidence, obesity has the potential to influence and alter these variables individually or collectively, contributing to the pathogenesis of type 2 diabetes (DeFronzo et al., 2015). A growing body of evidence indicates that body mass index is strongly associated with insulin resistance and diabetes mellitus. There are certain substances that are involved in the development of insulin resistance, such as non-esterified fatty acids, glycerol, hormones, cytokines, proinflammatory markers, etc. The amount of these substances is increased in obese individuals. The underlying cause of diabetes pathogenesis is the impairment of β-islet cells in the pancreas. β-islet cells of the pancreas are chiefly responsible for insulin and amylin production. This insulin plays its role in moving glucose in the liver and muscle cells. Contrarily, impaired pancreatic β-cells cause uncontrolled or abnormal blood glucose levels due to insufficient insulin production. When impaired pancreatic β-islet cells are accompanied by insulin resistance, the development of diabetes becomes unavoidable (Al-Goblan et al., 2014). Insulin resistance is defined as a condition when the liver, fat, and muscle cells do not respond to insulin and, consequently, cannot uptake glucose from the blood into the cells (National Institute of Diabetes and Digestive and Kidney Diseases, 2018). Insulin resistance is associated with body mass index to any extent of an increase in body weight.

Another point to consider regarding insulin sensitivity is fat distribution in the body. It differs in lean and obese individuals because of how fat is distributed in the body. Individuals who have fat accumulated towards the periphery are more insulin sensitive than those who have fat accumulated in the abdomen and chest areas (Karpe et al., 2011). This may happen because abdominal fat is more lipolytic than subcutaneous fat. Insulin has anti-lipolytic properties, and abdominal fat does not react to it easily. Hence, abdominal fat accumulation is considered a crucial factor in insulin resistance and diabetes pathogenesis (Roden et al., 1996; Fain et al., 2004).

Previous evidence also suggests that obese people in the general population, specifically men, suffer from sleep apnea (Li et al., 2010). Pharyngeal fat buildup is a significant risk factor for sleep apnea development, evident in 80% of Sleep AHEAD study participants (Foster et al., 2009).

Being overweight or obese significantly increases the likelihood of osteoarthritis development. Osteoarthritis most commonly occurs in the knee joint in overweight and obese patients but can also affect the hip joint, wrist joint, or any other joint as the pathophysiology includes an increase in body mass, the circulating adipokines, inflammatory factors, etc. All these factors result in osteoarthritis prevalence in obese and overweight individuals (Reyes et al., 2016).

Substantial weight gain markedly elevates the likelihood of developing hypertension and end-stage kidney disease (ESRD). Obesity raises blood pressure by promoting sodium retention, impairing regulation, and activating sympathetic nervous and renin-angiotensin systems (Hall, 2003). In response, the kidneys dilate and hyper-filter to uphold sodium equilibrium. However, with time, these adjustments, along with heightened arterial pressure, impose stress on the kidneys, leading to glomerular injury (Hall et al., 1993). Consequently, there may be an increase in urinary protein excretion, a gradual decline in nephron function, and exacerbated hypertension. Moreover, obese individuals with type II diabetes and metabolic disruptions may experience accelerated kidney disease progression (Hall, 2004).

Obesity is also associated with an increased risk of nonalcoholic fatty liver disease (NAFLD), which is characterized by steatosis (an increase in intrahepatic triglycerides). The prevalence of NAFLD is directly associated with increasing body mass index (BMI). An analysis of liver tissues obtained from several biopsies suggests that the prevalence percentages of steatosis and steatohepatitis (intra-hepatic fat accumulation associated with inflammation and fibrosis) are about 15% and 3%, respectively, in non-obese individuals, while 65% and 20%, respectively, in people with class I and II obesity (BMI 30.0–39.9 kg/m2), and 85% and 40%, respectively, in extremely obese patients (BMI ≥40 kg/m2) (Fabbrini et al., 2010).

Depression is also connected to an elevated risk of gaining weight and developing obesity, and these conditions, in turn, are associated with a greater vulnerability to depressive disorders. A recent study revealed that individuals with obesity have a 55% higher likelihood of experiencing lifetime depression, whereas those with depression are 58% more prone to becoming obese compared to the general population (Luppino et al., 2010).

1.4 Histological Background of Obesity Progression to an Epidemic

It has been a few decades since the obesity issue became an epidemic on a global level. After the eighteenth century, due to technological advancements, food supply increased, which aided in improved health outcomes and better quality, amount, and variety of food. This led to increased longevity and body sizes. However, as a consequence of the circumstances faced in the Second World War, these technological advancements led to excessive production of readily available processed foods, and a reduction in physical activity was noted that gradually led to an increase in the prevalence of obesity (Eknoyan, 2006).

Several studies have been conducted, and theories have been formulated to explain the altering food environment and its relevance to the obesity epidemic. It has been known that specific nutrients such as fats and carbohydrates have been theorized to be the prime cause of obesity in individuals, but these theories are not factual. Instead, the obesity epidemic has resulted because of the change in caloric density and quality of food supply as the industrialized food system revolutionized and increased its production and marketing of processed and ultra-processed foods

using cheap agricultural materials such as corn and soy. Such foods are inexpensive and conveniently available at corner stores. They are laden with excessive amounts of sodium, sugar, fat, flavor additives, and other ingredients that increase their craving and consumption. Marketing and high availability have drawn people's attention and led to increased consumption of these time-saving convenient foods (Hall, 2018).

Experimental studies on individual nutrients have been conducted to identify the accountability of these nutrients in the progression of the obesity epidemic. It was thought that the protein leverage model played a role in causing obesity, which was explained as when protein consumption is low, the body achieves the target intake by consuming non-protein caloric food sources. However, the fact that the protein composition of the food supply did not considerably decrease throughout an upsurge in obesity prevalence in the U.S. provides more evidence against the protein leverage model. Therefore, it appears that the protein leverage hypothesis of obesity does not adequately explain the global epidemic of obesity. Similarly, experimental studies support that increased fat consumption leads to fat accumulation in the body, increasing body fat mass and vice versa; however, the obesity prevalence did not regress by the widespread recommendations to adopt lower-fat diets throughout the 1980s and 1990s.

Furthermore, compared to other diets, low-fat prescription diets did not result in higher weight reduction among people (Tobias et al., 2015). On the other hand, it was thought that carbohydrates, especially simple or refined ones, contributed to the increased obesity prevalence by increasing caloric consumption from refined carbohydrate-rich foods in the food supply. This could be true based on the evidence that global increase in weight gain has been positively associated with an increase in per capita food energy availability, and the extent of this positive correlation is more than enough to explain the weight gain that has been reported in 80% of countries. This indicates that increased energy availability in the food supply is a decisive contributor to the obesity epidemic (Hall, 2018).

In 2022, a global cost of living crisis emerged due to surging energy and food prices. This has led to short-term financial struggles and long-term economic hardships for many families (Whiting, 2022). The inflation surge has reduced disposable income, impacting food purchasing, and may push economically disadvantaged families into extreme poverty and starvation. The crisis also intersects with the existing obesity problem, forcing families to choose between affordable but unhealthy food and costly nutritious options. Financial constraints make healthier choices harder for lower-income households, contributing to higher BMI (Robinson et al., 2022). The cost of living crisis may exacerbate socioeconomic disparities in obesity rates, disproportionately affecting vulnerable communities. Government responses will vary, but some may prioritize short-term economic growth over anti-obesity policies, as the U.K. did. However, this approach risks worsening obesity rates and causing long-term economic consequences. The current global obesity burden is substantial and likely to worsen if unaddressed (BBC News, 2022). To tackle the crisis effectively, governments must consider the broader implications of obesity within the context of other challenges like climate change and financial crises. Ignoring these connections may perpetuate the obesity crisis indefinitely (Robinson, 2022).

1.5 Previous Approaches & Limitations

It's essential to recognize that tackling obesity is a complex issue requiring a multifaceted and collaborative approach involving governments, healthcare systems, food industries, urban planners, educators, and individuals themselves. Previous strategies employed by various countries have been extensive in scope, all had each of their limitations. Several public health campaigns have been conducted in different states of the U.S. to educate the public on obesity and weight management by organizations such as those conducted by the Obesity Action Coalition, including Your Weight Matters Campaign, Stop Weight Bias campaign, Obesity care week (Obesity Action Coalition, n.d.). Such public health campaigns and education efforts have been valuable in raising awareness about obesity and its consequences. However, their impact may be limited by factors such as the target audience's receptiveness to the messages, the effectiveness of communication channels, and the need for sustained efforts to achieve significant behavior change. One such example is Australia's national obesity campaign, Measure Up, which was launched in 2008. It aimed to

raise awareness of obesity-related health risks and encourage healthy behaviors. Evaluations showed increased awareness and knowledge among people but limited impact on self-reported physical activity and eating habits. Another phase, Swap It, Don't Stop It, promoted specific healthy swaps, but only 16% of Australians made changes. Researchers suggested additional environmental changes to support behavior change. Mass media campaigns were useful in increasing awareness but may need complementary strategies to achieve significant and sustained behavioral improvements (Obesity Evidence Hub, n.d.a).

Several countries have established specific dietary guidelines to promote adequate food and nutrient consumption using cultural foods. However, these guidelines are for professionals who can comprehend them well but not for ordinary people. Because of this reason, the public remains unaware of healthy portion sizes of foods and keeps on following the fluctuating trends of the modern world, such as the consumption of fast foods, junk foods, and processed foods consisting of higher amounts of added sugars, sodium and saturated fats per serving. Moreover, implementing and enforcing food policies produced using these guidelines can be challenging due to various factors, including the influence of the food industry, cultural preferences, and individual choices.

Some countries applied food taxation to prevent people from buying and consuming many unhealthy foods. Hungary introduced taxes on foods with high fat, sugar, salt, and caffeine content. The government also increased its tariffs on carbonated beverages and alcohol. The tax amount varied depending on the goods. The additional charge, for instance, on soft drinks was €0.016 per liter, prepackaged sweetened goods cost €0.33 more per kilogram, salty snacks cost an additional €0.67 per kilogram, and energy drinks cost €0.84 more per liter.

Similarly, Denmark targeted saturated fats in foods by implementing a levy of €2.41 per kilogram of saturated fat used in the production of a particular food that surpasses 2.3% of saturated fat content resulting in a higher tax on fatty foods such as hamburgers. Likewise, France incorporated sugar-sweetened beverages taxes (Villanueva, 2011). While imposing taxes on unhealthy foods and beverages can discourage their consumption, such a move may also face opposition from the food industry and may not be well-received by consumers. Additionally, it may disproportionately affect lower-income populations and lead to potential unintended consequences.

Healthcare interventions for obesity, such as medical treatments and weight loss programs, can be effective but may not be accessible to everyone due to cost and healthcare disparities. According to an article published by Cornell University, obesity accounts for a direct medical care cost of $289 billion annually, increasing costs in all categories of care, including inpatient hospital stays, outpatient doctor visits, and prescription drugs (Hanchett, 2022). Long-term adherence to treatment programs can also be challenging for some individuals.

Several countries have introduced clear and easy-to-understand nutrition labels on food packaging to help consumers make informed choices about their food purchases. These labels often highlight critical nutritional information, such as calories, sugar, fat, and salt content. Some countries have implemented regulations on food advertising, especially targeting unhealthy foods and beverages, particularly those aimed at children. This approach seeks to reduce the influence of marketing on food choices and promote healthier options. In 2016, Chile was the first country to implement front-of-package mandatory labeling to limit certain nutrients in packaged foods. Chile requires a warning label in the form of a black stop sign on packaged foods and beverages that exceed sugar, salt, saturated fat, or calorie limits. For each nutrient exceeded, products must include a 'stop' warning sign, which means some products can have up to four labels. Marketing of food aimed at children and food sold in preschools and schools is also restricted by law. This means that foods with a stop sign warning must not be marketed or sold to children. After Chile, several countries, including Peru, Uruguay, and Israel, have implemented nutrient warning labels on processed foods and beverages to address health concerns related to excessive salt, sugar, fats, and saturated fats.

In Peru and Uruguay, black and white octagonal labels indicate products exceeding nutrient thresholds. In Israel, red warning labels with text and images are used for high-sugar, salt, and saturated fat products, with stricter criteria phased in gradually. Mexico also approved black and white octagonal labels for excess sugar, sodium, fats, and calories. In addition, products with warning labels cannot target children through characters or celebrities. Colombia introduced warning labels in 2022, while Brazil, Argentina, Canada, and South Africa are actively considering similar measures. However, implementing and enforcing these regulations may face industry opposition and require

careful monitoring. Moreover, a certain level of education is needed to accurately apprehend the messages coming from the signage. For instance, traffic sign labels with green (depicting frequent consumption of a certain nutrient), yellow (depicting moderate consumption of a certain nutrient), and red (depicting occasional consumption of a certain nutrient) colors may be interpreted differently by some people (Obesity Evidence Hub, n.d.b).

2. Materials and Methods

2.1 Data Mining - Analyzing Large Databases to Generate New Information

Data mining, also known as "knowledge discovery in data," is the act of identifying patterns and other important information in massive data sets. Given the rise of large data, data mining techniques have expanded in recent decades, supporting various organizations, including research institutes, in translating raw data from target data sets into valuable knowledge. These methods are helpful in organizing and filtering the data, surfacing the most interesting information, and then describing the data set to analyze it, producing valuable information. Scientists explain this data using observations of patterns, associations, and correlations (IBM, n.d.).

Statistical methods such as linear regression are one data-mining technique that can help analyze country-wide obesity indicators to identify their contribution to obesity prevalence among different nations. It enables the researchers to detect a relationship between obesity indicators, including obesity rates, BMI, waist circumference, and potential influencing factors such as income, education level, physical activity, access to healthcare, and dietary patterns. In this way, researchers can understand which factors are associated with higher or lower obesity rates in different countries. Moreover, linear regression can be used to generate predictive models to estimate obesity rates in countries according to relevant variables. Such models can be beneficial for projecting obesity trends and evaluating the potential impact of interventions or policy changes. Country-wise obesity data analysis through linear regression can be helpful in comparing the magnitude and direction (positive or negative; direct or indirect) of associations between obesity rates and its indicators from different regions of the world.

Recognizing the factors that drive obesity in different nations can help guide the development and execution of specific programs and policies to combat the problem. For example, if linear regression analysis in a study indicates that in a particular nation, poor physical activity is substantially related to high obesity rates, policymakers might focus on encouraging physical activity efforts in that region. Following the implementation of interventions and policies to counter the rising obesity prevalence, linear regression may be used to assess their effectiveness in addressing the issue. Researchers can investigate if the implementation of specific interventions impacts changes in obesity markers, assisting in evaluating the effectiveness of the measures used. The statistical analysis of obesity indicators aids scientific knowledge of the intricate interactions between numerous contributing factors and obesity rates. This knowledge can be used to develop more targeted and practical strategies to combat obesity globally. For instance, in a study conducted by Ramasamy et al., obesity indicators, including waist circumference, hip circumference, waist-to-height ratio, waist-to-hip ratio, and body mass index (BMI), were analyzed to find associations with health-related quality of life (HRQOL) using correlation coefficient (R) and coefficient of determination (R^2) to identify the strength of the associations, the direction (direct or inverse) and the extent to which these anthropometric measures contribute to the variability in the data. Results indicated that higher BMI is correlated with lower scores for HRQOL in middle-aged South Indian women who are morbidly obese (Ramasamy et al., 2020). Another study utilized multiple linear regression models to find associations between Resilience Scale, Mood Survey, emotional stress, and stress control scores with waist circumference and BMI among healthy older adults in Great Britain and Portugal. The goal of the study was to determine the extent to which the psychological factors contributed to the variations in the BMI and waist circumference data. The study suggested that the models for obesity indicators in Great Britain offered a more robust explanation than in Portugal, potentially due to cultural differences in understanding psychosocial, lifestyle, and psychological factors in obesity etiology. Both countries indicated the need to target males, less educated individuals, and

to promote resilience. In Great Britain, efforts should also focus on healthcare providers and the unemployed, aiming to reduce alcohol consumption and increase physical activity. Additionally, the study highlighted the significance of researching positive psychological factors to enhance understanding of obesity (Stewart-Knox et al., 2012).

Overall, using statistical methods like linear regression in obesity research improves our understanding of this global health issue, helps identify risk factors as well as protective factors, and informs evidence-based policies and interventions to address obesity in diverse populations.

2.2 Linear Regression

Linear regression seeks to model the relationship between two variables by fitting a linear equation from the observed data set. In a simple linear regression model, there are two variables, i.e., an independent or explanatory variable and a dependent variable. For instance, a researcher might want to find the relationship of the weights of individuals to their heights through a linear regression model. A linear regression line takes the form of an equation; $Y = a + bx$, where 'x' is the independent variable and 'Y' is the dependent variable (Linear Regression, n.d.). The dependent variable is also known as outcome variable, criterion variable, endogenous variable, or regression. The independent variable is also called the exogenous variable, predictor variable, or regressor (Statistics Solutions, 2013). 'b' is the slope of the line or the magnitude which x impacts y, and 'a' is the line intercept (which is the value of y when $x = 0$). There are many names for a regression's dependent variable (Linear Regression, n.d.).

A researcher or statistician should first evaluate whether or not there is a link between the variables of interest before attempting to fit a linear model to the observed data set. This does not always indicate that one variable is the cause of the other variable (for instance, higher SAT scores do not necessarily result in better college grades) but rather that there is a substantial relationship or association between the two variables. To fulfill this purpose, a scatterplot can be used to assess the strength of the relationship between two variables. If there appears to be no relationship between the explanatory and dependent variables, i.e., no increasing or decreasing trends in the scatterplot are seen, fitting a linear regression model to the data is unlikely to provide a meaningful model (Linear Regression, n.d.).

Linear regression is a fundamental and widely used kind of predictive analysis. The overall goal of regression is to investigate two things: (1) How well does a collection of predictor factors predict an outcome (dependent) variable? (2) Which factors are significant predictors of the outcome variable, and how do they influence it (as indicated by the size and sign of the beta estimates)? Simply, researchers use regression analysis to determine the strength of predicting variables, projecting an effect, and trend estimation (Statistics Solutions, 2013).

2.3 Types of Linear Regression

Linear regression is not limited to simple functioning. Based on the technicality of the model and type of variables used, there are several kinds of linear regression models, including simple, multiple, logistic, ordinal, multinomial, and discriminant regression analysis (Statistics Solutions, 2013).

1. Simple Linear Regression

This model analyzes the association between one dependent variable, coded using an interval or ratio scale, and one independent variable, with data following an interval, ratio, or dichotomous scale.

2. Multiple Linear Regression

Multiple linear regression controls for multiple covariates or independent variables to predict the outcome. This model analyzes the relationship between one dependent variable, coded using an interval or ratio scale, and two or more independent variables, with data following an interval, ratio, or dichotomous scale.

3. *Logistic Regression*

Logistic regression analyzes the relationship between one dependent variable, coded using a dichotomous scale, and two or more independent variables, with data following interval, ratio, or dichotomous scale. In other words, it is used when the target variable is binary or categorical, meaning the outcome can take only two possible values, such as "yes" or "no," "0" or "1," "True" or "False," etc.

4. *Ordinal Regression*

An ordinal regression model involves the association analysis between one dependent variable, coded using an ordinal scale, and two or more independent variables, with data following a nominal or dichotomous scale. It is also known as ordinal logistic regression and comprises a dependent variable that has an inherent order or ranking among its categories, but the differences between the categories may not be uniform or quantifiable. It is used when the dependent variable can take on three or more ordered categories, often represented by integer values or ordered labels. Examples of ordinal variables include education level (e.g., high school, college, graduate), customer satisfaction levels (e.g., very dissatisfied, neutral, very satisfied), or BMI levels (e.g., underweight, normal, overweight, obese).

5. *Multinomial Regression*

A multinomial regression model or multinomial logistic regression is a type of regression analysis used when the dependent variable has three or more discrete and unordered categories. In multinomial regression, the goal is to model the probabilities of an observation belonging to each category of the dependent variable, given the input features (independent variables). It is particularly useful when dealing with nominal or categorical outcome variables. The model estimates a set of coefficients for each feature corresponding to each possible outcome category.

6. *Discriminant Analysis*

A discriminant analysis model involves the association analysis between one dependent variable, coded using a nominal scale, and one or more independent variables, with data following an interval, ratio scale. Multiple linear regression is usually more desirable by researchers or statisticians for data analysis because it allows them to estimate or predict the outcome changes when multiple factors change simultaneously. The estimated coefficients from a linear regression analysis estimate the magnitude of change in the dependent variable by a one-unit increase or decrease in an independent variable and are reported with 95% confidence intervals and a probability value (p-value) which tests the hypothesis of whether the covariate is associated with the outcome (Kulaylat, 2023). Exemplifying this in terms of the paper, we need to analyze the obesity predictors and their relationship with obesity rates. So, using a multiple linear regression model will aid in identifying which factors are contributors to the increasing prevalence of obesity in various nations of the world.

2.4 Coefficient of Determination – R^2

The coefficient of determination or R^2 in linear regression determines the proportion of variation in data points described by the regression line. It describes how closely connected the data points are in the fitted line. The value of R^2 represents the amount of variation in the data, which is explained by the model or intrinsic factors or independent variables and tells how much variation in the data is because of external factors. It is calculated as the ratio of the overall variation of data points explained by the regression line (Sum of squared regression) to the total variance of data points from the mean (also known as the total sum of squares). The ratio is represented by the formula mentioned below. The prediction or a point on the regression line is denoted by y_hat, the mean of all the values is denoted by y_bar, and the actual values are represented by y_i (Kumar, 2022).

R^2 = (sum squared regression or SSR) / (total sum of squares or SST)
$R^2 = \sum(yi - \hat{y}i)^2 / \sum(yi - \bar{y})^2$.

It can also be computed as a function of the total variation of data points from the regression line (also known as the sum of square residual) and the total variation of data points from the mean.

R^2 =1− (sum squared regression or SSR) / (total sum of squares or SST)
$R^2 = 1 - \sum(yi - \hat{y}i)^2 / \sum(yi - \bar{y})^2$.

After we construct a multilinear regression model, there is a need to evaluate its performance. The model predictability performance can be measured through the coefficient of determination. However, adjusted R^2 is a more appropriate measure (Kumar, 2022). While representing the regression model graphically, there are two axes, i.e., the x-axis and y-axis. Various data points are plotted on the graph, and a regression line is built through the model. Assume that this regression line is diagonal and just around the plotted points, extended to the point where it intercepts the x-axis. There is another horizontal line in the middle of the plot that intersects the regression line as well. This horizontal line indicates the mean of all the values of the regression model's response variable. The vertical distance of a specific data point from the regression line is called the sum of squares residual error or SSE. It explains the total deviation of actual values from the regression line and represents the prediction error in terms of the sum of the squared distance between the actual values and the regression line's prediction. Similarly, the vertical distance of a certain point of the regression line to the horizontal line that acts as the mean of all values is called the sum of squares regression or SSR and describes the whole variation of prediction from the mean. When we combine both the vertical distances, we get the sum of squares total or SST. It is also known as the total error at times. It describes the variation of data values from the mean and is determined through the summation of the squared distances of each point from the mean.

A few key elements of R^2 should be recognized to understand its application in identifying the best-fit line or regression model. First, a higher value of SSR represents a better regression line. This also means that the value of SSR is closer to SST. As R^2 = SSR / SST, this implies that the value of R^2 is closer to 1. Second, the regression line is better when SSE has a lower value. This means that the regression line gets better when the value of SSE is closer to zero. As R^2= 1 - (SSE/SST), this implies that the value of R-squared is closer to 1 (Kumar, 2022).

2.5 Correctness

Correctness in linear regression refers to the extent to which the model's predictions align with the actual observed data. In other words, it assesses how well the linear regression model fits the data points and how accurately it can make predictions for new, unseen data. In linear regression, the objective is to find the best-fitting line that minimizes the difference between the predicted values and the actual target values. This difference is often measured using a metric called the "residual" or "error." The goal is to minimize the sum of squared residuals, a method known as the "Least Squares" approach.

Residuals aid in the identification of deviated observed values from the expected ones. A residual indicates how good our model is in comparison to the actual value through statistical and graphical representations (*How to Interpret R-squared in Regression Analysis?*, n.d.). While conducting linear regression analysis, a key assumption is that the residuals are normally distributed. While the linear model is resistant to deviations from normality, there can be instances, such as the presence of outliers, when it is better to alter the dependent variable before running the statistical analysis, such as, by taking the natural log. This transformation may enhance the model's fitness (Kulaylat, 2023). Moreover, prior to assessing the coefficient of determination or R^2, the residual plots should be determined. This is an essential step in creating a regression model. By spotting problematic trends in the residual plots, they aid in the identification of a biased model. However, one cannot rely on the outcomes if the model is biased. The value of

R^2 and other numerical outputs can be evaluated if the residual plots appear good (*How to Interpret R-squared in Regression Analysis?*, n.d.).

When choosing a model for the study, an important aspect to consider is model fitting, also known as 'goodness-of-fit,' which describes the difference between observed and expected values in a regression model. A common idea is that if the deviations among the observed and the predicted values of the linear model are small and unbiased (not reaching extreme values such as too high or too low), it is a well-fit model (*How to Interpret R-squared in Regression Analysis?*, n.d.). Another crucial factor to consider for model fitness is balancing the regression equation with a certain number of variables for accurate analysis outcomes. Introducing independent variables into a linear regression model will always raise the model's explained variance (R^2). Thus, when using too many variables, overfitting can occur, reducing the model's generalizability. If a model has a high number of variables, some of them will be statistically significant owing to chance alone. Occam's razor perfectly illustrates the problem: a simple model is typically preferable to a more complicated model. Because of such unignorable issues, it is recommended to be careful and only add a reasonable number of variables while not overcrowding the model (Statistics Solutions, 2013).

2.6 Paper Design and Data Collection

Obesity is a cross-sectional ecological study. The paper utilizes quantitative data and a literature review to gather previous and current insights into the global overweight and obesity trends. We have used secondary data from authentic and credible sources, including peer-reviewed articles, to analyze and predict the degree of obesity using linear regression, one of the data mining techniques, according to the development or welfare level of the country using various indicators by country such as socioeconomic status, physical activity, dietary habits and types of foods consumed, triglycerides, ALT (SGPT), glycated hemoglobin (HbA1c), and uric acid (Jeon et al., 2022; Dinsa et al., 2012). The results thus produced will be utilized to propose relevant policies that can be implemented on a national level to alleviate the current state and prevalence of obesity worldwide, as the contribution of obesity-predicting factors can vary among countries, and thus, policy recommendations might be different for each country.

3. Results

3.1. Table

Country	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expe	Hepatitis B	Measles	BMI	under-five death	Polio	Total expenditur	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 ye	thinness 5-9 yea	Income composi	Schooling	obesity
Afghanistan	59.9	271	64	0.01	73.52	62	492	18.6	86	58	8.18	62	0.1	612.7	327582	17.5	17.5	0.48	10	4.9
Albania	77.5	8	0	4.51	428.75	98	0	57.2	1	98	5.88	98	0.1	4575.76	288914	1.2	1.3	0.76	14.2	20.5
Algeria	75.4	11	21	0.01	54.24	95	0	58.4	24	95	7.21	95	0.1	547.85	39113313	6	5.8	0.74	14.4	26
Angola	51.7	348	67	8.33	23.97	64	11699	22.7	101	68	3.31	64	2	479.31	2692466	8.5	8.3	0.53	11.4	7.5
Argentina	76.2	118	8	7.93	847.37	94	1	62.2	9	92	4.79	94	0.1	12245.26	42981515	1	0.9	0.83	17.3	27.3
Armenia	74.6	12	1	3.91	295.61	93	13	54.1	1	95	4.48	93	0.1	3994.71	29622	2.1	2.1	0.74	12.7	19.2
Australia	82.7	6	1	9.71	10769.36	91	340	66.1	1	92	9.42	92	0.1	62214.69	2346694	0.6	0.6	0.94	20.4	27.9
Austria	81.4	66	0	12.32	8350.19	98	117	57.1	0	98	11.21	98	0.1	51322.64	8541575	1.8	2	0.89	15.9	19.2
Azerbaijan	72.5	119	5	0.01	306.18	94	0	51.5	6	97	6.4	94	0.1	7891.3	953579	2.8	2.9	0.75	12.2	18.8
Bangladesh	71.4	132	98	0.01	10.45	97	289	17.7	121	97	2.82	97	0.1	184.57	15945279	18.1	18.6	0.57	10	3.2
Belarus	72	199	0	13.94	1147.11	97	64	61.7	0	97	5.69	97	0.1	8318.43	9474511	1.9	2	0.8	15.7	23.7

3.2. Example of Factors

1. Preprocessing of the Data (Refining)

f = open('factor.csv', 'r', encoding='utf-8')
rdr = csv.reader(f)
d = dict()
flag =0
firstline = []
for line in rdr:
if flag == 0:
firstline = line
if flag == 1:
for l in range(1,len(line)):
line[l] = float(line[l])
d[line[0]] = line[1:]
flag = 1
	Generate a dictionary by the country and convert the string to float value.

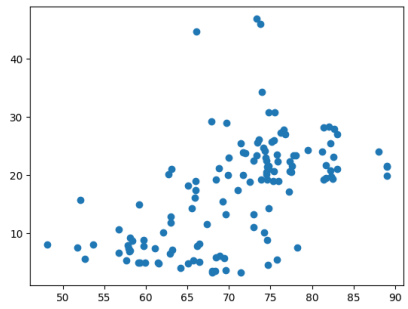
f1 = open('ob.csv', 'r', encoding='utf-8')
f2 = open('data.csv', 'w', encoding='utf-8')
rdr1 = csv.reader(f1)
rdr2 = csv.writer(f2)
for line in rdr1:
if line[0] in d:
temp = [line[0]]
for dd in d[line[0]]:
temp.append(dd)
for i in range(len(line[1])):
if line[1][i] == '[':
line[1] = line[1][:i-1]
break
temp.append(line[1])
rdr2.writerow(temp)
	Generate the 'data.csv' file which contains the whole data for analysis.

f = open('data.csv', 'r', encoding='utf-8') rdr = csv.reader(f) for line in rdr: print(line)	0 seconds ['\ufeffCountry', 'Life expectancy ', 'Adult Mortality', 'infant deaths', 'Alcohol', 'percentage expenditure', 'Hepatitis B', 'Measles ', 'BMI ', 'under-five deaths ', 'Polio', 'Total expenditure', 'Diphtheria ',' HIV/AIDS', 'GDP', 'Population',' thinness 1-19 years',' thinness 5-9 years', 'Income composition of resources', 'Schooling', 'obesity'] ['Afghanistan', '59.9', '271.0', '64.0', '0.01', '73.52', '62.0', '492.0', '18.6', '86.0', '58.0', '8.18', '62.0', '0.1', '612.7', '327582.0', '17.5', '17.5', '0.48', '10.0', '4.9'] ['Albania', '77.5', '8.0', '0.0', '4.51', '428.75', '98.0', '0.0', '57.2', '1.0', '98.0', '5.88', '98.0', '0.1', '4575.76', '288914.0', '1.2', '1.3', '0.76', '14.2', '20.5'] ['Algeria', '75.4', '11.0', '21.0', '0.01', '54.24', '95.0', '0.0', '58.4', '24.0', '95.0', '7.21', '95.0', '0.1', '547.85', '39113313.0', '6.0', '5.8', '0.74', '14.4', '26.0'] ['Angola', '51.7', '348.0', '67.0', '8.33', '23.97', '64.0', '11699.0', '22.7', '101.0', '68.0', '3.31', '64.0', '2.0', '479.31', '2692466.0', '8.5', '8.3', '0.53', '11.4', '7.5'] ['Argentina', '76.2', '118.0', '8.0', '7.93', '847.37', '94.0', '1.0', '62.2', '9.0', '92.0', '4.79', '94.0', '0.1', '12245.26', '42981515.0', '1.0', '0.9', '0.83', '17.3', '27.3']

2. Generating the Model

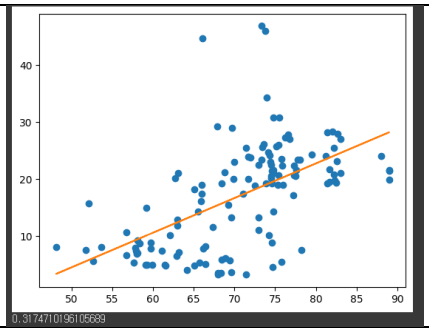
```
from sklearn.linear_model import LinearRegression
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv("data.csv")
print(df.head())

X = df["Life expectancy "]
y = df["obesity"]
plt.plot(X, y, 'o')
plt.show()
line_fitter = LinearRegression()
line_fitter.fit(X.values.reshape(-1,1), y)
```

```
       Country  Life expectancy   Adult Mortality  infant deaths  Alcohol  ₩
0  Afghanistan              59.9            271.0           64.0     0.01
1      Albania              77.5              8.0            0.0     4.51
2      Algeria              75.4             11.0           21.0     0.01
3       Angola              51.7            348.0           67.0     8.33
4    Argentina              76.2            118.0            8.0     7.93

   percentage expenditure  Hepatitis B  Measles    BMI   under-five deaths   ₩
0                   73.52         62.0     492.0  18.6                 86.0
1                  428.75         98.0       0.0  57.2                  1.0
2                   54.24         95.0       0.0  58.4                 24.0
3                   23.97         64.0   11699.0  22.7                101.0
4                  847.37         94.0       1.0  62.2                  9.0

    ...  Total expenditure  Diphtheria    HIV/AIDS      GDP  Population  ₩
0   ...               8.18         62.0       0.1   612.70    327582.0
1   ...               5.88         98.0       0.1  4575.76    288914.0
2   ...               7.21         95.0       0.1   547.85  39113313.0
3   ...               3.31         64.0       2.0   479.31   2692466.0
4   ...               4.79         94.0       0.1  12245.26  42981515.0

    thinness  1-19 years   thinness 5-9 years  ₩
0                   17.5                 17.5
1                    1.2                  1.3
2                    6.0                  5.8
3                    8.5                  8.3
4                    1.0                  0.9

   Income composition of resources  Schooling  obesity
0                             0.48       10.0      4.9
1                             0.76       14.2     20.5
2                             0.74       14.4     26.0
3                             0.53       11.4      7.5
4                             0.83       17.3     27.3
```



3. Evaluating the Model

```
plt.plot(X, y, 'o')
plt.plot(X,line_fitter.predict(X.values.reshape(-1,1)))
plt.show()
print(line_fitter.score(X.values.reshape(-1,1), y))
```



3.3 Evaluation

1. Single Factor

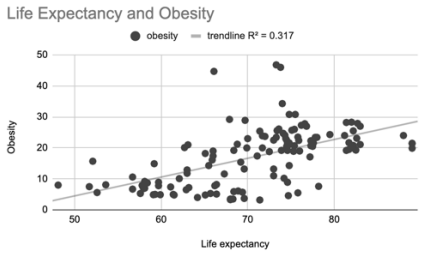


Figure 1.

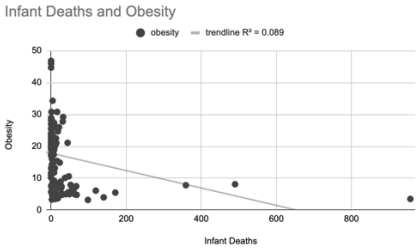


Figure 2.

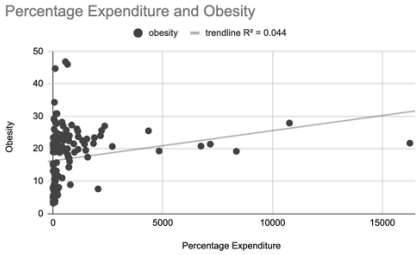


Figure 3.

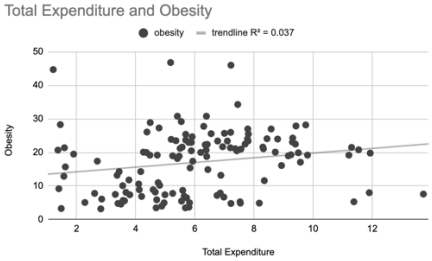


Figure 4.

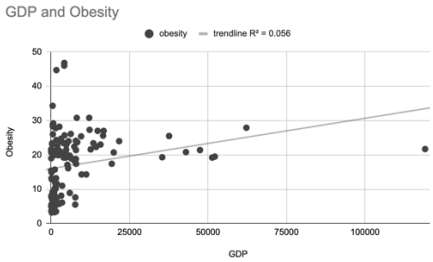


Figure 5.

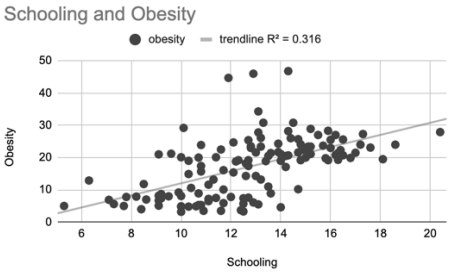


Figure 6.

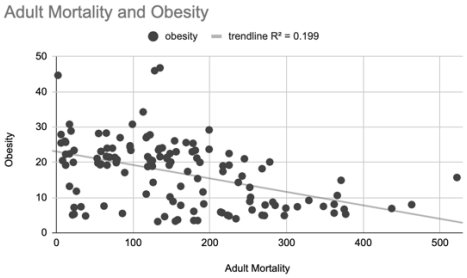


Figure 7.

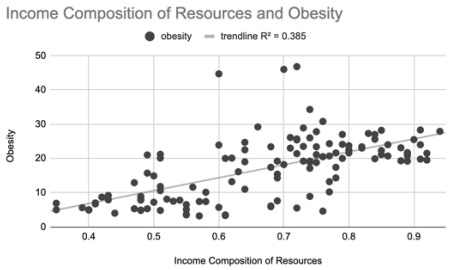


Figure 8.

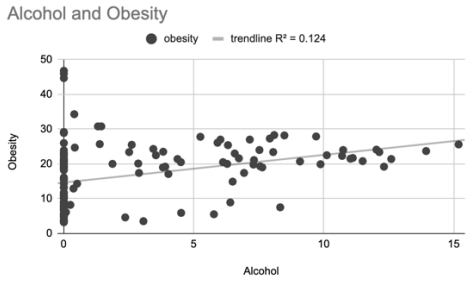
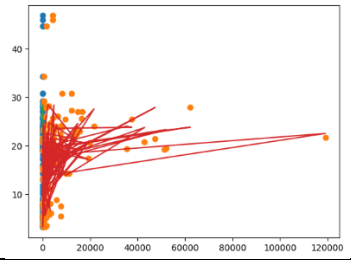
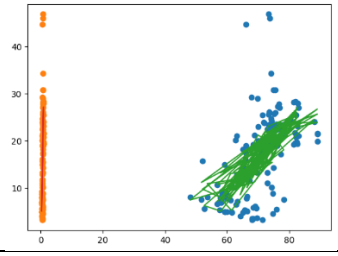
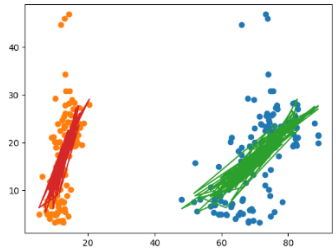


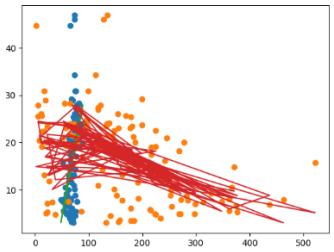
Figure 9.

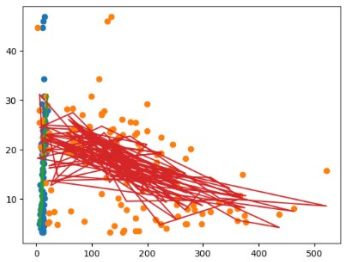
2. Multifactor

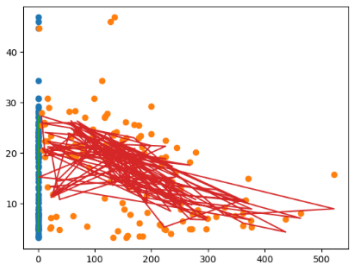
Factor: X = df[["Life expectancy ", "GDP"]]
R^2: 0.31776343332178836 

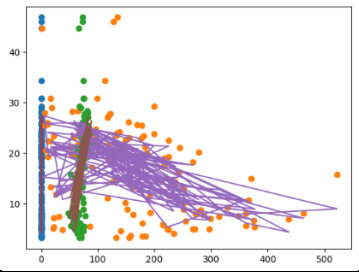
Factor: X = df[["Life expectancy, ""Income composition of resources"]]
R^2: 0.31776343332178836 

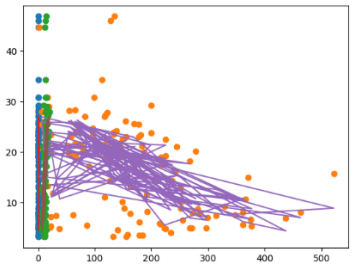
Factor: X = df[["Life expectancy ", "Schooling"]]
R^2: 0.3503889833340671 

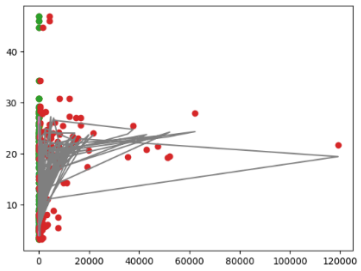
Factor: X = df[["Life expectancy ", "Adult Mortality"]]
R^2: 0.31776343332178836 

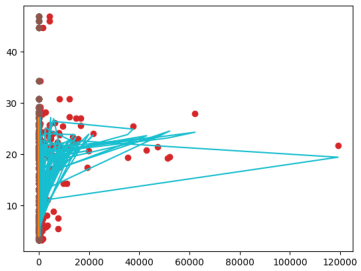
Factor: df[["Schooling," "Adult Mortality"]]
R^2: 0.33744476016187175 

Factor: X = df[["Income composition of resources", "Adult Mortality"]]
R^2: 0.3862207247703675 

Factor: X = df[["Income composition of resources," "Adult Mortality," "Life expectancy"]]
R^2: 0.3862207249523364 

Factor: X = df[["Income composition of resources," "Adult Mortality," "Schooling"]]
R^2: 0.3863731373259469 

Factor: X = df[["Income composition of resources," "Adult Mortality," "Schooling," "GDP"]]
R^2: 0.3935569885792538 

Factor: X = df[["Income composition of resources," "Adult Mortality," "Schooling," "GDP," "Life expectancy, "Alcohol"]]
R^2: 0.39369710692424753 

4. Discussion

In this paper, I discussed the numerous factors that increase obesity nationally and globally. I evaluated multiple factors to analyze the correlation between these factors and obesity in each country using linear regression. Collecting the data to join and preprocessing them are also contribution points of this paper. The results demonstrated that life expectancy, schooling level, and income composition of resources strongly correlated with the target value in the single factor aspect. In contrast, national alcohol consumption, GDP level, and others are comparatively weak. In the multi-factor evaluation, the composition of income, adult mortality, schooling, GDP, life expectancy, and national alcohol intake had the highest correlation with the target value. Thus, governments should focus on not only improving a single factor, such as schooling or life expectancy, but multiple factors simultaneously.

References

Al-Goblan, A. S., Al-Alfi, M. A., & Khan, M. Z. (2014). Mechanism linking diabetes mellitus and obesity. *Diabetes, Metabolic Syndrome and Obesity : Targets and Therapy*, *7*, 587–591. https://doi.org/10.2147/DMSO.S67400

BBC News. (2022). Anti-obesity strategy to be reviewed due to cost-of-living crisis. https://www.bbc.co.uk/news/uk-politics-62900076.

Behl, S., & Misra, A. (2017). Management of obesity in adult Asian Indians. *Indian Heart Journal*, *69*(4), 539–544. https://doi.org/10.1016/j.ihj.2017.04.015

Chen, Y., Peng, Q., Yang, Y., Zheng, S., Wang, Y., & Lu, W. (2019). The prevalence and increasing trends of overweight, general obesity, and abdominal obesity among Chinese adults: a repeated cross-sectional study. *BMC Public Health*, *19*(1), 1293. https://doi.org/10.1186/s12889-019-7633-0

DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., Hu, F. B., Kahn, C. R., Raz, I., Shulman, G. I., Simonson, D. C., Testa, M. A., & Weiss, R. (2015). Type 2 diabetes mellitus. Nature reviews. Disease primers, 1, 15019. https://doi.org/10.1038/nrdp.2015.19

Dhurandhar, N. V., Petersen, K. S., & Webster, C. (2021). Key causes and contributors of obesity: a perspective. *Nursing Clinics*, *56*(4), 449-464. https://www.nursing.theclinics.com/article/S0029-6465(21)00070-0/fulltext

Dhurandhar, N.V. What is obesity?. *Int J Obes* 46, 1081–1082 (2022). https://doi.org/10.1038/s41366-022-01088-1

Dinsa, G. D., Goryakin, Y., Fumagalli, E., & Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: a systematic review. *Obesity Reviews*, *13*(11), 1067–1079. https://doi.org/10.1111/j.1467-789X.2012.01017.x

Duffey, K. J., Gordon-Larsen, P., Jacobs, D. R., Jr, Williams, O. D., & Popkin, B. M. (2007). Differential associations of fast food and restaurant food consumption with 3-y change in body mass index: the Coronary Artery Risk Development in Young Adults Study. The American journal of clinical nutrition, 85(1), 201–208. https://doi.org/10.1093/ajcn/85.1.201

Eknoyan, G. (2006). A history of obesity, or how what was good became ugly and then bad. *Advances in Chronic Kidney Disease*, *13*(4), 421–427. https://doi.org/10.1053/j.ackd.2006.07.002

Fabbrini, E., Sullivan, S., & Klein, S. (2010). Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology*, *51*(2), 679–689. https://doi.org/10.1002/hep.23280

Fain, J. N., Madan, A. K., Hiler, M. L., Cheema, P., & Bahouth, S. W. (2004). Comparison of the release of adipokines by adipose tissue, adipose tissue matrix, and adipocytes from visceral and subcutaneous abdominal adipose tissues of obese humans. *Endocrinology*, *145*(5), 2273–2282. https://doi.org/10.1210/en.2003-1336

Foster, G. D., Sanders, M. H., Millman, R., Zammit, G., Borradaile, K. E., Newman, A. B., Wadden, T. A., Kelley, D., Wing, R. R., Sunyer, F. X., Darcey, V., Kuna, S. T., & Sleep AHEAD Research Group (2009). Obstructive sleep apnea among obese patients with type 2 diabetes. Diabetes care, 32(6), 1017–1019. https://doi.org/10.2337/dc08-1776

Fox, A., Feng, W., & Asal, V. (2019). What is driving global obesity trends? Globalization or "modernization"? *Globalization and Health*, *15*(1), 32. https://doi.org/10.1186/s12992-019-0457-y

Garcia-Mantrana, I., & Collado, M. C. (2016). Obesity and overweight: Impact on maternal and milk microbiome and their role for infant health and nutrition. Molecular nutrition & food research, 60(8), 1865–1875. https://doi.org/10.1002/mnfr.201501018

Hall, J. E. (2003). The kidney, hypertension, and obesity. *Hypertension*, *41*(3), 625-633.

Hall, J. E., Brands, M. W., Dixon, W. N., & Smith Jr, M. J. (1993). Obesity-induced hypertension. Renal function and systemic hemodynamics. Hypertension, 22(3), 292-299.

Hall, J. E., Henegar, J. R., Dwyer, T. M., Liu, J., Da Silva, A. A., Kuo, J. J., & Tallam, L. (2004). Is obesity a major cause of chronic kidney disease?. *Advances in renal replacement therapy*, *11*(1), 41-54. https://www.sciencedirect.com/science/article/pii/S1073444903000669

Hall, K. D. (2018). Did the food environment cause the obesity epidemic? *Obesity*, *26*(1), 11–13. https://doi.org/10.1002/oby.22073

Hanchett, J. (December 20, 2022). How much money is too much for obesity treatments?. Cornell Chronicle. Retrieved July 30, 2023, from https://news.cornell.edu/stories/2022/12/how-much-money-too-much-obesity-treatments

Hou, X., Jia, W., Bao, Y., Lu, H., Jiang, S., Zuo, Y., Gu, H., & Xiang, K. (2008). Risk factors for overweight and obesity, and changes in body mass index of Chinese adults in Shanghai. *BMC Public Health*, *8*, 389. https://doi.org/10.1186/1471-2458-8-389

How to Interpret R-squared in Regression Analysis? (n.d.). Knowledge hut. Retrieved August 6, 2023, from https://www.knowledgehut.com/blog/data-science/interpret-r-squared-and-goodness-fit-regression-analysis#what-is-regression-analysis?%C2%A0

IBM. (n.d.). What is Data Mining? Retrieved August 4, 2023, from https://www.ibm.com/topics/data-mining

Jeon, J., Lee, S., & Oh, C. (2022). Age-specific risk factors for the prediction of obesity using a machine learning approach. *Frontiers in Public Health*, *10*, 998782. https://doi.org/10.3389/fpubh.2022.998782

Jia, W. (2015). Obesity in China: its characteristics, diagnostic criteria, and implications. *Frontiers of Medicine*, *9*(2), 129–133. https://doi.org/10.1007/s11684-015-0387-x

Karpe, F., Dickmann, J. R., & Frayn, K. N. (2011). Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes*, *60*(10), 2441–2449. https://doi.org/10.2337/db11-0425

Kelly, T., Yang, W., Chen, C. S., Reynolds, K., & He, J. (2008). Global burden of obesity in 2005 and projections to 2030. *International Journal of Obesity*, *32*(9), 1431–1437. https://doi.org/10.1038/ijo.2008.102

Klein, S., Gastaldelli, A., Yki-Järvinen, H., & Scherer, P. E. (2022). Why does obesity cause diabetes? *Cell Metabolism*, *34*(1), 11–20. https://doi.org/10.1016/j.cmet.2021.12.012

Kulaylat, A. N., Tran, L., Kulaylat, A. S., & Hollenbeak, C. S. (2023). Regression analysis. In *Translational Surgery* (pp. 157–163). Elsevier. https://doi.org/10.1016/B978-0-323-90300-4.00087-2

Kumar, A. (February 21, 2022). R-squared, R2 in linear regression: Concepts, examples. Data Analytics. Retrieved August 6, 2023, from https://vitalflux.com/r-squared-explained-machine-learning/#

Leibel, R. L., Rosenbaum, M., & Hirsch, J. (1995). Changes in energy expenditure resulting from altered body weight. New England Journal of Medicine, 332(10), 621-628. https://www.nejm.org/doi/full/10.1056/nejm199503093321001

Li, C., Ford, E. S., Zhao, G., Croft, J. B., Balluz, L. S., & Mokdad, A. H. (2010). Prevalence of self-reported clinically diagnosed sleep apnea according to obesity status in men and women: National Health and Nutrition Examination Survey, 2005-2006. Preventive medicine, 51(1), 18–23. https://doi.org/10.1016/j.ypmed.2010.03.016

Lin, X., & Li, H. (2021). Obesity: Epidemiology, Pathophysiology, and Therapeutics. Frontiers in endocrinology, 12, 706978. https://doi.org/10.3389/fendo.2021.706978

Linear Regression. (n.d.). Yale. Retrieved August 6, 2023, from http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Penninx, B. W., & Zitman, F. G. (2010). Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Archives of general psychiatry, 67(3), 220-229. https://jamanetwork.com/journals/jamapsychiatry/article-abstract/210608

National Institute of Diabetes and Digestive and Kidney Diseases. (May, 2018). Insulin Resistance & Prediabetes. NIDDK. Retrieved August 4, 2023, from https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance#

Obesity Action Coalition. (n.d.). Awareness. Retrieved July 30, 2023, from https://www.obesityaction.org/about/awareness/

Obesity Evidence Hub. (n.d.a). Overweight and obesity mass media education campaigns. Retrieved July 30, 2023, from https://www.obesityevidencehub.org.au/collections/environmental/mass-media-education-campaigns

Obesity Evidence Hub. (n.d.b). Nutrient warning labels. Retrieved July 30, 2023, from https://www.obesityevidencehub.org.au/collections/prevention/nutrient-warning-labels

Obri, A., & Claret, M. (2019). The role of epigenetics in hypothalamic energy balance control: implications for obesity. Cell stress, 3(7), 208.

Ramasamy, S., Joseph, M., Jiwanmall, S. A., Kattula, D., Nandyal, M. B., Abraham, V., Samarasam, I., Paravathareddy, S., Paul, T. V., Rajaratnam, S., Thomas, N., & Kapoor, N. (2020). Obesity Indicators and Health-related Quality of Life - Insights from a Cohort of Morbidly Obese, Middle-aged South Indian Women. *European Endocrinology*, *16*(2), 148–151. https://doi.org/10.17925/EE.2020.16.2.148

Reyes, C., Leyland, K. M., Peat, G., Cooper, C., Arden, N. K., & Prieto-Alhambra, D. (2016). Association Between Overweight and Obesity and Risk of Clinically Diagnosed Knee, Hip, and Hand Osteoarthritis: A Population-Based Cohort Study. Arthritis & rheumatology (Hoboken, N.J.), 68(8), 1869–1875. https://doi.org/10.1002/art.39707

Robinson, E. (2022). Obesity and the cost of living crisis. *International Journal of Obesity*, 1–2. https://doi.org/10.1038/s41366-022-01242-9

Robinson, E., Jones, A., & Marty, L. (2022). The role of health-based food choice motives in explaining the relationship between lower socioeconomic position and higher BMI in U.K. and U.S. adults. *International Journal of Obesity*, *46*(10), 1818–1824. https://doi.org/10.1038/s41366-022-01190-4

Roden, M., Price, T. B., Perseghin, G., Petersen, K. F., Rothman, D. L., Cline, G. W., & Shulman, G. I. (1996). Mechanism of free fatty acid-induced insulin resistance in humans. *The Journal of Clinical Investigation*, *97*(12), 2859–2865. https://doi.org/10.1172/JCI118742

Schnabel, R. B., Yin, X., Larson, M. G., Yamamoto, J. F., Fontes, J. D., Kathiresan, S., Rong, J., Levy, D., Keaney, J. F., Jr, Wang, T. J., Murabito, J. M., Vasan, R. S., & Benjamin, E. J. (2013). Multiple inflammatory biomarkers in relation to cardiovascular events and mortality in the community. Arteriosclerosis, thrombosis, and vascular biology, 33(7), 1728–1733. https://doi.org/10.1161/ATVBAHA.112.301174

Statistics Solutions. (2013). What is Linear Regression . Retrieved on August 6th, 2023, from https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/

Stewart-Knox, B., E Duffy, M., Bunting, B., Parr, H., Vas de Almeida, M. D., & Gibney, M. (2012). Associations between obesity (BMI and waist circumference) and socio-demographic factors, physical activity, dietary habits, life events, resilience, mood, perceived stress and hopelessness in healthy older Europeans. *BMC Public Health*, *12*, 424. https://doi.org/10.1186/1471-2458-12-424

Tan, K. C. B. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*. https://hub.hku.hk/handle/10722/77434

Tobias, D. K., Chen, M., Manson, J. E., Ludwig, D. S., Willett, W., & Hu, F. B. (2015). Effect of low-fat diet interventions versus other diet interventions on long-term weight change in adults: a systematic review and meta-analysis. *The Lancet Diabetes & Endocrinology*, *3*(12), 968-979.

Villanueva, T. (2011). European nations launch tax attack on unhealthy foods. *Canadian Medical Association Journal*, *183*(17), E1229-30. https://doi.org/10.1503/cmaj.109-4031

Wariri, O., Alhassan, J. A. K., Mark, G., Adesiyan, O., & Hanson, L. (2021). Trends in obesity by socioeconomic status among non-pregnant women aged 15-49 y: a cross-sectional, multi-dimensional equity analysis of demographic and health surveys in 11 sub-Saharan Africa countries, 1994-2015. *International Health*, *13*(5), 436–445. https://doi.org/10.1093/inthealth/ihaa093

Whiting, K. (September 21, 2022). The cost-of-living crisis is having a global impact. Here's what countries are doing to help. World economic forum. Retrieved on July 31st, from https://www.weforum.org/agenda/2022/09/cost-of-living-crisis-global-impact/

World Health Organization. (June 9th, 2021). Obesity and overweight. Retrieved July 30, 2023, from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

World Health Organization. (n.d). Obesity. Retrieved July 30, 2023, from https://www.who.int/health-topics/obesity#tab=tab_1

Yao, Y., Cai, X., Chen, C., Fang, H., Zhao, Y., Fei, W., ... & Zheng, C. (2020). The role of microbiomes in pregnant women and offspring: research progress of recent years. *Frontiers in Pharmacology*, *11*, 643. https://www.frontiersin.org/articles/10.3389/fphar.2020.00643/full

Yao, Y., Cai, X., Ye, Y., Wang, F., Chen, F., & Zheng, C. (2021). The Role of Microbiota in Infant Health: From Early Life to Adulthood. Frontiers in immunology, 12, 708472. https://doi.org/10.3389/fimmu.2021.708472

Yatsuya, H., Folsom, A. R., Yamagishi, K., North, K. E., Brancati, F. L., Stevens, J., & Atherosclerosis Risk in Communities Study Investigators (2010). Race- and sex-specific associations of obesity measures with ischemic stroke incidence in the Atherosclerosis Risk in Communities (ARIC) study. Stroke, 41(3), 417–425. https://doi.org/10.1161/STROKEAHA.109.566299

Yoo, S. (2018). Dynamic energy balance and obesity prevention. Journal of obesity & metabolic syndrome, 27(4), 203.