Analyis of Undernourishment Predictors and Responses Throughout the World

Final Report

Team 3: Alayna Binder, Hanzhi Peng, Karly Purkey, Jingsong Shan

December 9, 2022

Introduction and Data

Malnutrition, defined by the World Health Organization (WHO) as deficiencies, excesses, or imbalances in a person's intake of energy and/or nutrients, is a significant challenge among countries worldwide (*Malnutrition*, 2020). In this project, we will specifically investigate undernutrition among child and adult populations, which is encompassed by stunting (low height for age), wasting (low weight for height), underweight, and micronutrient deficiencies (a lack of important vitamins and minerals). It is vital to note that overweight and obesity also fall under malnutrition; 462 million adults worldwide are underweight, while a staggering 1.9 billion are considered overweight. Children with severe malnutrition have an increased risk of serious illness and death due to acute infectious diseases (Bhutta et al., 2017). Out of the total 209 million children impacted by stunting or wasting, those under the age of 5 living in low- and middle-income countries are predominantly affected.

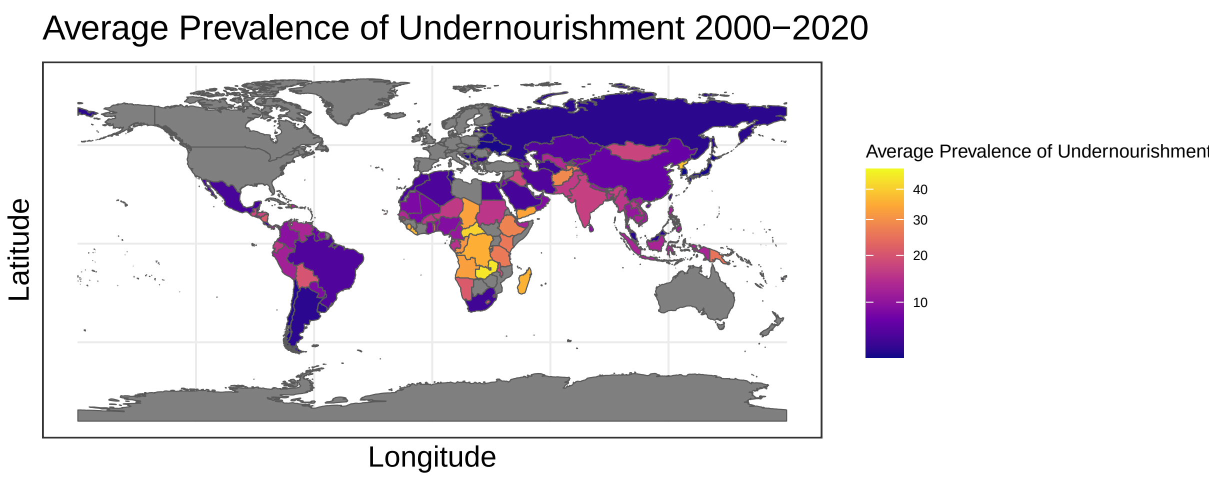
By analyzing the potential contributors to global variation in malnutrition trends and its complications, we seek to answer the following research questions: (1) How do trends in food security and their association with nutritional health relate to malnutrition rates in countries worldwide? (2) Which long-term physical complications may be associated with malnutrition, and how can we best address those, given the connection between food security and nutritional health?

Our data set, "Suite of Food Security Indicators," is derived from the Food and Agriculture Organization of the United Nations (FAO), which presents the core set of food security indicators. Field experts established these indicators based on availability, access, utilization, and stability–the four pillars of food security. The data was collected by the FAO Statistics Division (ESS) beginning in 1999 and ranging through 2022, with certain variables calculated on a 3-year average consisting of 204 countries.

We hypothesized that the potential contributors to malnutrition were related to food security and nutritional needs and therefore selected variables related to (1) food supply: "Average fat supply (g/cap/day) (3-year average)", "Average protein supply (g/cap/day) (3-year average)", "Average dietary energy supply adequacy (percent) (3-year average)", "Dietary energy supply used in the estimation of the prevalence of undernourishment (kcal/cap/day) (3-year average)", food security: "Prevalence of severe food insecurity in the total population (percent) (3-year average)", "Prevalence of moderate or severe food insecurity in the total population (percent) (3-year average)", and (2) nutrition requirements: "Average dietary energy requirement (kcal/cap/day)", "Minimum dietary energy requirement (kcal/cap/day)" to examine the correlation between these variables and the severity of malnutrition. Because our data sets included population-level statistics for each country, the data collection was complicated and required compiling multiple data sources for each variable. For the majority of the data sets, surveys were used to collect samples then models were produced to obtain estimates for the entire population. The models often included multiple parameters that were accurate predictors of the variable. The data collection methods for the prevalence of undernourishment were highly complex and varied based on the country and the data available to the researchers. The general method for estimating the prevalence of undernourishment in a population was to use the habitual dietary intake levels for a population and the minimum dietary energy requirement, dependent on sex and age.

According to the World Bank, the countries are classified into four groups by income to investigate the impact of developmental status on malnutrition. Utilizing the database "Food Prices for Nutrition" in the World Bank, we also explored the role of health diet access in malnutrition, including variables "Percent of the population who cannot afford a healthy diet at 52% of income" and "Cost of a Healthy Diet." The cost of a healthy diet is defined as a diet that provides enough calories as well as sufficient levels of essential nutrients needed to maintain a healthy life, outlined by scientifically supported national food-based dietary guidelines. The cost of a healthy diet is calculated by composing a diet that meets these requirements and uses the least expensive items available in each country. The percent of the population who cannot afford a healthy diet at 52% of income was obtained by comparing the cost of a healthy diet to the income distributions for each country. According to the World Bank, 52% of income was used as the cutoff, because, at the international poverty line, 52% of income is spent on food.

The world map below illustrates the average prevalence of undernourishment by percentage in countries worldwide between the years 2000-2020.



In the context of this project, a high prevalence of undernourishment is greater than or equal to 10%. Compared to other continents, Africa–especially countries in the central and south region–has the highest global prevalence of undernourishment. This is followed by southeast and south Asia, as well as northwest South America. Russia, China, and countries in East Europe and the Middle East do not have a high prevalence of undernourishment.

Data Cleaning and Data Wrangling

To tidy the data, the first step was to convert the values into doubles and remove all rows that were "NA". Because some of the yearly values in the FAO derived data set were calculated by taking the average over three years, we had to mutate the names of the years to be single year doubles rather than character string ranges. For example, to get the prevalence of undernourishment in one country in 2020, the data collectors took the average prevalence of undernourishment from 2019, 2020, and 2021. In the original data set, the year was “2019-2021”, and we mutated it to be 2020. We did this to all observations that were collected as an average from 3 years in order to analyze the data over time and compare variables that were yearly estimates with variables that were obtained from three-year averages. It should be noted that there was no exact value for countries with prevalence of undernourishment under 2.5% (since it was indicated by <2.5%). So when transforming the data, these values were omitted since they would not have a large impact on our research. While it is possible to find data on some of the missing countries that have internal reporting on prevalence of undernourishment, they would (1) have a different methodology for data collection (i.e. by year instead of a three-year average), which makes comparisons less accurate, and (2) lack many of the other specific variable categories found in the FAO data set and therefore not prove beneficial to our further analyses. Additionally, in order to merge the FAO data set with other data sets, we had to create extra variables that matched the country names in each of the three data sets, since initially there were inconsistencies with country names. For example in the FAO data set, the US was listed as “United States of America”, but in the other two data sets it was listed as the “United States of America”. Finally, the tidy FAO data set was merged with the data set obtained from the WHO by country name in order to perform data analyses between the variables in the two data sets. The joint FAO and WHO data set was then merged with a separate data set that paired countries with world region and a code that enabled us to create data visualization on a map.

Summary Tables To summarize our data we generated the following summary tables to show the average values for each of our variables by region.

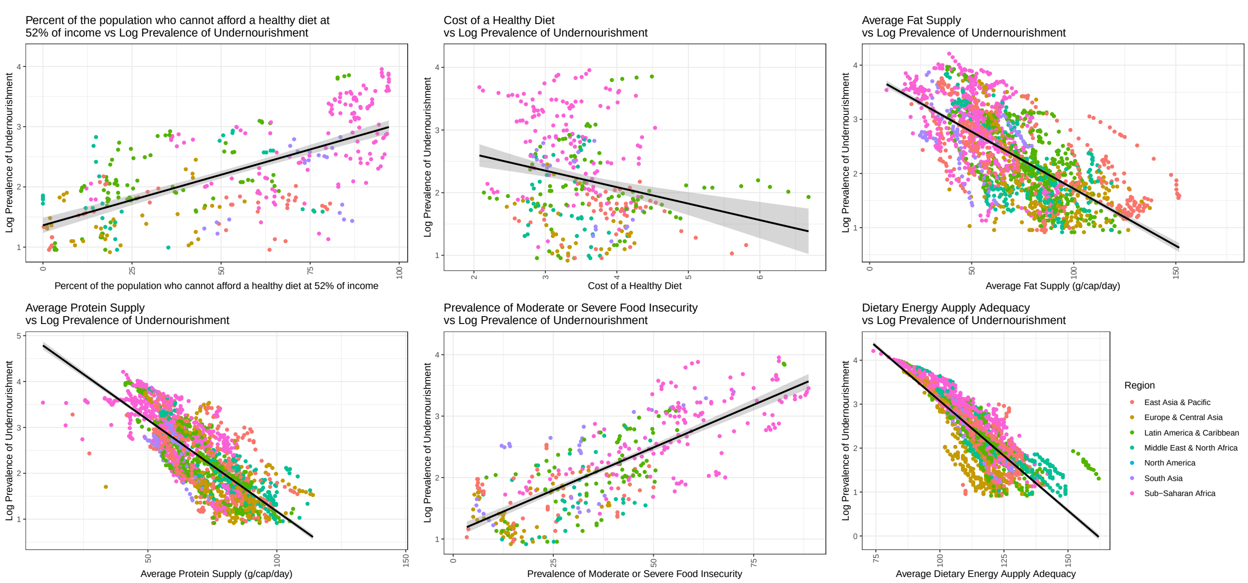
Table 1: Summary of Potential Contributing Factors Related to Prevalence of Undernourishment

Region	Average Prevalence of Undernourishment	Average Cost of a Healthy Diet	Average Percent of the Population who Cannot Afford a Healthy Diet at 52% of Income	Average Fat Supply (g/cap/day)	Average Protein Supply (g/cap/day)	Average Prevalence of Severe Food Insecurity in the Total Population (%)	Average Prevalence of Moderate or Severe Food Insecurity in the Total Population (%)	Average Dietary Energy Supply Adequacy (percent)
East Asia & Pacific	12.281425	3.739716	31.508197	86.58241	77.69664	4.6884615	18.51977	118.0017
Europe & Central Asia	8.126523	3.037477	6.468605	114.65072	96.02864	2.1446903	10.43105	127.7117
Latin America & Caribbean	11.621004	3.827301	30.907865	78.53693	73.01648	10.8333333	32.34865	118.0379
Middle East & North Africa	10.399609	3.251703	26.897727	85.72013	86.68021	8.0507692	24.19692	127.7812
North America	NaN	3.344500	1.162500	143.53478	107.44722	0.8083333	7.20000	143.8500
South Asia	15.959167	3.667571	59.450000	48.03889	58.94815	8.5903226	28.99032	108.5500
Sub-Saharan Africa	21.192429	3.298183	76.611765	52.50014	60.99338	24.1573034	58.07978	109.0629

Table 2: Summary of Potential Complications Related to Prevalence of Undernourishment

Region	Average Prevalence of Undernourishment	Prevalence of Obesity in the Adult Population	Prevalence of Low Birthweight (%)	Prevalence of Anemia Among Women of Reproductive Age	Percentage of Children Under 5 Years of Age who are Stunted (%)	Percentage of Children Under 5 years of Age who are Overweight (%)	Percentage of Children Under 5 years Affected by Wasting (%)
East Asia & Pacific	12.281425	19.540784	10.344118	25.93726	22.285538	5.433157	7.109184
Europe & Central Asia	8.126523	18.728799	6.143484	18.26638	10.054213	10.148535	4.430137
Latin America & Caribbean	11.621004	19.389965	10.315278	21.07129	15.226701	7.639626	2.591803
Middle East & North Africa	10.399609	24.572059	10.234896	29.90000	17.288235	10.568067	6.022951
North America	NaN	27.938235	7.021875	9.25250	2.861905	9.428571	0.420000
South Asia	15.959167	3.791912	19.657500	41.26562	35.945833	3.394643	12.442500
Sub-Saharan Africa	21.192429	7.198632	15.103333	41.79511	33.873016	5.073268	7.869328

Potential Contributors to Malnutrition



Linear Regression Model with Multiple Predictors

For Potential Contributors to Malnutrition

Table 3: Coefficient-Level Estimates for a Model Fitted to Estimate Factors Related to Prevalence of Undernourishment

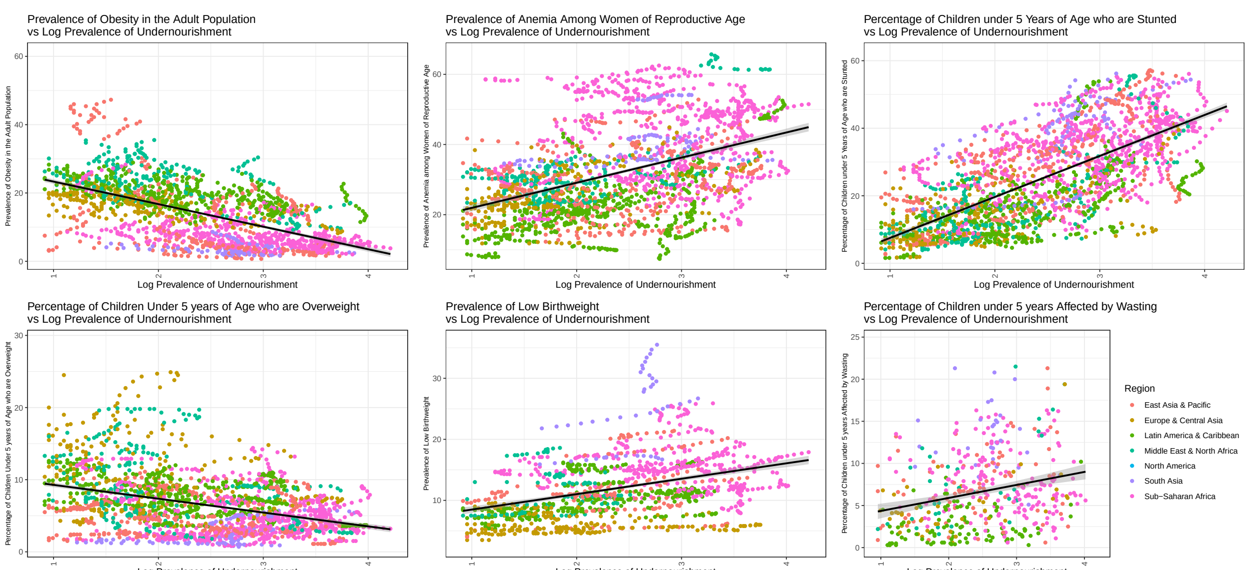
Predictor	B	SE	t	p
(Intercept)	5.36	0.253	21.13	0
‘Percent of the population who cannot afford a healthy diet at 52% of income [CoHD_headcount]‘	0.01	0.001	7.02	0
‘Prevalence of moderate or severe food insecurity in the total population (percent) (3-year average)‘	0.01	0.001	10.51	0
‘Average dietary energy supply adequacy (percent) (3-year average)‘	-0.03	0.002	-17.37	0
‘Cost of a healthy diet [CoHD]‘	-0.12	0.028	-4.21	0

Table 4: R-squared Value for a Model Fitted to Estimate Factors Related to Prevalence of Undernourishment

R-squared
0.8556

This model predicts the log transformed prevalence of undernourishment for a country from the percent of the population who cannot afford a healthy diet at 52% of income, the prevalence of moderate or severe food insecurity in the total population, the average dietary energy supply adequacy, and the cost of a healthy diet. Multiple models were generated using various factors within the data set, and these four variables were found to be the strongest predictors. This was determined by the p-values of the model as well as the r-squared value (0.86), which indicates that 86% of the variance for prevalence of undernourishment is explained by the model. The predicted variable was log transformed, because the original distribution had a strong skew. These coefficient estimate for these four variables had p-values less than 0.05 which indicates there is a significant relationship between these predictors and the prevalence of undernourishment. The model estimates that with all else held constant, for each increase in percent of the population who cannot afford a healthy diet increases, we would expect the log prevalence of undernourishment to be higher, on average, by 0.0061 (95% CI: 0.0044 - 0.078). With all else held constant, for every increase of the prevalence of moderate or severe food insecurity, the estimate increases by 0.014 (95% CI: 0.011 - 0.017) on average. With all else held constant, for every increase of the percent dietary supply adequacy, the estimate decreases by 0.031 (95% CI: -0.034 - -0.027) on average. With all else held constant, the estimate decreases by 0.12 (95% CI: -0.17 - -0.063) for every dollar increase in the calculated cost of a healthy diet.

Potential Complications of Malnutrition



Linear Regression Models with from a Single Predictor

For Potential Complications of Malnutrition

Wasting and Stunted Development in Children Under 5

As defined by UNICEF, wasting is low weight-for-height and is the "most visible and lethal type of malnutrition. Severe wasting, also known as severe acute malnutrition, is its most deadly form." Children who are stunted have been prevented from growing or developing properly, most directly caused by inadequate nutrition.

Table 5: Coefficient-Level Estimates for a Model Fitted to Estimate Relation of Log Percentage of Children Under 5 who are Affected by Wasting and Stunting to Log Prevalence of Undernourishment

Predictor	B	SE	t	p
(Intercept)	2.04	0.080	25.42	0
log('Prevalence of undernourishment (percent) (3-year average)')	0.54	0.031	17.20	0

Table 6: R-squared Value for a Model Fitted to Estimate Relation of Log Percentage of Children Under 5 who are Affected by Wasting and Stunting to Log Prevalence of Undernourishment

R-squared
0.3824

According to the adjusted R-squared value, ~38% of the variance in the percentage of children under 5 affected by wasting or are stunted in this regression model can be explained by the prevalence of undernourishment.

Overweight in Children Under 5

The definition for overweight in children used in the FAO data, as set by the World Health Organization, is weight-for-height greater than 2 standard deviations above WHO Child Growth Standards median.

Table 7: Coefficient-Level Estimates for a Model Fitted to Estimate Relation of Log Percentage of Overweight Children Under 5 Related to Log Prevalence of Undernourishment

Predictor	B	SE	t	p
(Intercept)	2.32	0.044	52.85	0
log('Prevalence of undernourishment (percent) (3-year average)')	-0.27	0.017	-15.55	0

Table 8: R-squared Value for a Model Fitted to Estimate Relation of Log Percentage of Overweight Children Under 5 to Log Prevalence of Undernourishment

R-squared
0.0996

According to the adjusted R-squared value, ~10% of the variance in the percentage of children under 5 overweight in this regression model can be explained by the prevalence of undernourishment.

Obesity in Adult Population

For adults, obesity is a BMI greater than or equal to 30, where BMI is a person's weight in kilograms divided by the square of their height in meters.

Table 9: Coefficient-Level Estimates for a Model Fitted to Estimate Log Relation of Prevalence of Obese Adults to Log Prevalence of Undernourishment

Predictor	B	SE	t	p
(Intercept)	3.72	0.049	75.32	0
log(‘Prevalence of undernourishment (percent) (3-year average)‘)	-0.55	0.020	-28.36	0

Table 10: R-squared Value for a Model Fitted to Estimate Relation of Log Prevalence of Obese Adults to Log Prevalence of Undernourishment

R-squared
0.3036

According to the adjusted R-squared value, ~30% of the variance in the prevalence of obese adults in this regression model can be explained by the prevalence of undernourishment.

Low Birthweight

Low birth weight is defined as a birth weight of less than 2500 g, as per the World Health Organization.

Table 11: Coefficient-Level Estimates for a Model Fitted to Estimate Relation of Log Prevalence of Low Birthweight to Log Prevalence of Undernourishment

Predictor	B	SE	t	p
(Intercept)	1.84	0.036	50.44	0
log(‘Prevalence of undernourishment (percent) (3-year average)‘)	0.23	0.015	15.81	0

Table 12: R-squared Value for a Model Fitted to Estimate Relation of Log Prevalence of Low Birthweight to Log Prevalence of Undernourishment

R-squared
0.1623

According to the adjusted R-squared value, ~16% of the variance in the prevalence of low birthweight in this regression model can be explained by the prevalence of undernourishment.

Anemia Among Women of Reproductive Age

Women of reproductive age is defined as 15-49 years old.

Table 13: Coefficient-Level Estimates for a Model Fitted to Estimate Relation of Log Prevalence of Anemia Among Women of Reproductive Age to Log Prevalence of Undernourishment

Predictor	B	SE	t	p
(Intercept)	2.82	0.026	107.40	0
log(‘Prevalence of undernourishment (percent) (3-year average)‘)	0.24	0.010	22.43	0

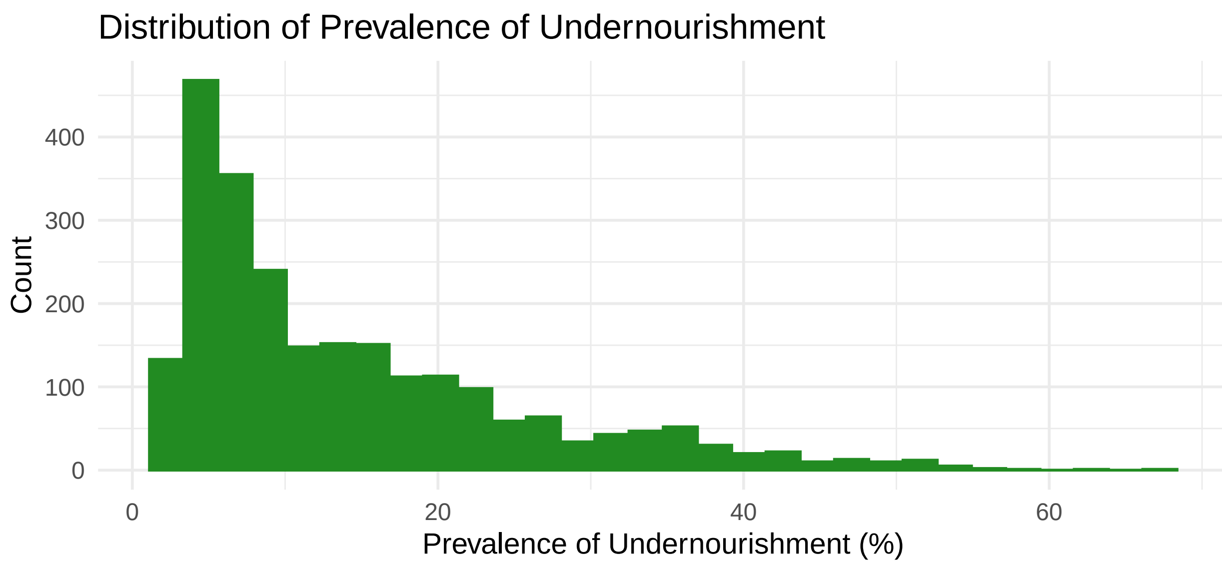
Table 14: R-squared Value for a Model Fitted to Estimate Relation of Log Prevalence of Anemia Among Women of Reproductive Age to Prevalence of Undernourishment

R-squared
0.1854

According to the adjusted R-squared value, ~19% of the variance in the prevalence of anemia among women of reproductive age in this regression model can be explained by the prevalence of undernourishment.

Chosen Methods

We considered two potential models in our research: ANOVA and linear regression. A graph of the distribution of prevalence of undernourishment was firstly made to determine whether ANOVA was appropriate in answering our research questions. As shown in the graph below, the data distribution is not a normal distribution but right skewed, which indicates that an ANOVA test would not be appropriate. To better address the research questions, the linear regression model was used for the following reasons: first, a linear regression model can be used to predict future outcomes. Based on the linear regression model we made, we can predict the future malnutrition rate based on the given predictors and thus provide suggestions on how to lower the prevalence of malnutrition. Second, whether the variables in the data set are statistically significant can be determined by the regression analysis. Third, the confidence interval can be arranged based on how precise we want our models to be, and the variables to include in the model can also be determined accordingly, which helps us adjust our model. Finally, the linear regression model provides a R squared value or adjusted R squared value, which provides evidence on how well our model addresses the research questions.



Based on the linear regression table, all the terms that were chosen to fit model has a p.value less than 0.01, and the R square calculated was 0.8556189. This indicates that all the variables that were chosen to fit in the model are statistically significant, and the model is able to explain 85.6% percent of the variance of the prevalence of undernourishment. Since the majority of the variance can be explained by the model, the research question is thus answered effectively.

Limitations

In our data analysis, we started by separating the variables into two categories ((1) contributors to malnutrition and (2) complications of malnutrition) in order to answer both of our research questions. For some of these variables, however, it was hard to decide whether the variable could have potentially influenced malnutrition rates, or, conversely, whether malnutrition rates could have potentially influenced the variable. For instance, the average dietary energy supply adequacy was categorized as a contributor to prevalence of undernourishment (i.e. low average dietary energy supply leads to malnutrition), but it can also be seen as a consequence of malnutrition (since malnutrition may result in low average dietary energy supply). This limitation would affect the linear regression models created and the variance in the prevalence of undernourishment they could explain.

A drawback of the FAO data set was that it had lots of missing data. Many of these variables were specific and thus likely difficult to measure among countries with lower levels of health reporting. As described previously, although it would be possible to individually track down internal data from countries with good health reporting, the method in which the data were measured could vary between the country's collection and that of the FAO. Also, the FAO data set was broad in that it included a wide variety of variables–finding each variable (for each country, for many different year) so that it could be fit in the regression model would be incredibly time consuming and challenging.

Discussion

One of the goals for this data exploration was to determine what factors have the strongest relationship with rates of undernourishment in order to identify potential solutions. From our analysis it was found that the cost of a healthy diet, the percent of the population who cannot afford a healthy diet, the dietary energy supply adequacy, and the prevalence of moderate or severe food insecurity. Furthermore, in response to our second research question, we determined that wasting and stunting, low birthweight, and anemia among women of reproductive age all increased with the prevalence of undernourishment (whereas percentage of overweight children under 5 and prevalence of obese adults expectedly decreased with the prevalence of undernourishment).

By developing policies and initiatives that target these specific factors, the global prevalence of undernourishment and its consequences can be reduced. The first step is to promote economic growth in countries with undernourishment and ensure that additional income earned by those impoverished/struggling with malnutrition is allocated towards improving both the quantity and quality of their health and diets. Governments can likewise combat malnutrition by allotting any additional public funds towards public goods and services, particularly in the interest of benefitting those who lack secure access to food. While there is a long way to go before undernourishment is eliminated, implementing programs focused on social protection and equitable access to resources by the poor, as well as holding institutions accountable for their transparency (or lack thereof) and protecting human rights, is a step in the positive direction.

References

- Bhutta, Z. A., Berkley, J. A., Bandsma, R. H., Kerac, M., Trehan, I., & Briend, A. (2017). Severe childhood malnutrition. *Nature Reviews Disease Primers*, 3(1). https://doi.org/10.1038/nrdp.2017.67
- *Malnutrition.* (2020, April 15). World Health Organization. https://www.who.int/news-room/questions-and-answers/item/malnutrition
- *FAOSTAT Suite of Food Security Indicators.* (2022, November 7). Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data/FS
- *DataBank Food Prices for Nutrition.* (2022, July 5). The World Bank. https://databank.worldbank.org/source/food-prices-for-nutrition