Death from the Womb: Uncovering links between Maternal, Malnutrtion, and Neonatal Deaths

Team Avatar: Xavier Ghazaleh, Abigail Phillips, Saajan Patel, Harriet Wiser

November 15, 2020

INTRODUCTION

Death is a certainty, but what we die of depends on the conditions (environmental, economic, political, etc.) under which we live. A significant factor that determines these conditions is our position in the world, both geographically and socially. This means that causes of death and their prevalence can vary based on our country and our region. This data set puts numbers to those disparities. Understanding how someone's situation affects their risks is broadly important because it determines what interventions are put in place, governmental or otherwise.

With this dataset, we would like to specifically explore the links between malnutrition, neonatal, and maternal deaths. These issues are linked to multiple UN Sustainable Development goals, including Gender Equality and Good Health and Well-Being. Additionally, both the time period after being born and the time period after giving birth have been some of the most dangerous time-periods for human beings historically. However, in recent decades, this has been improving in some locations - particularly high-income ones. Malnutrition is another issue that is more common in low-income countries; we therefore wonder if the rates of death from nutritional deficiencies are linked to the rates of maternal and neonatal mortality (there is also existing evidence for this being the case). While we probably cannot determine causation with just this data, even correlation is interesting as it points towards the existence of some set of phenomena that may worsen these conditions.

The numbers are drawn from a study called the Global Burden of Disease done by the medical journal The Lancet. It is a world-wide observational study that has collected data starting in 1990 and updated up to 2017, although we are using a version that ends in 2016. The data we analyze includes mortality rates by the percent of deaths caused by 32 issues every year in 228 regions/countries. The original study describes the data as being collected through "censuses, household surveys, civil registration and vital statistics, disease registries, health service use, air pollution monitors, satellite imaging, disease notifications, and other sources" and the percentages calculated using their own statistical methods including "the Cause of Death Ensemble model and spatiotemporal Gaussian process regression."

The causes of death (observations) seem to be roughly organized by prevalence, although this, of course, varies. The first three are cardiovascular disease, cancer, and respiratory disease. The countries and regions (variables) are ordered alphabetically, with each location having all 27 years they were tracked represented right after each other (i.e. it is ordered by location not by timeline). The first country is Afghanistan, and the fifth location gives us our first region, Andean Latin America.

In short, our general research question is: how are maternal mortality rates related to the deaths rates of nutritional deficiencies, neonatal deaths, and protein-energy malnutrition?

METHODOLOGY

General variables used within the data analysis:

- `country`: Region of deaths being analyzed.
- `Maternal deaths (%)`: Percent of deaths in each region due to maternal issues.
- `Neonatal deaths (%)`: Percent of deaths in each region due to neonatal issues.
- `Nutritional deficiencies (%)` : Percent of deaths in each region due to nutritional deficiencies.
- `Protein-energy malnutrition (%)`: Percent of deaths in each region due to protein-energy malnutrition.

To first explore the research question, preliminary data analysis was collected on regional maternal and neonatal death rates for each region within the dataset in the year 2016. To accomplish this a `global_2016_reg` variable was created to filter only for maternal death rates and neonatal death rates for specific regions around the world in the year 2016. This variable purposely excludes individual countries as we are trying to understand a broader geographical relationship. A bar chart was used to graph the categorical variables, which are the regions, against the numerical percent death rate.

However, for the main analysis, we did want to include country-by-country information. Thus, the `global_2016` was created to filter for deaths rates in the year 2016 for every country in the dataset. Neonatal death rates, nutritional deficiency death rates, and protein-energy malnutrition death rates, were plotted against maternal death rates.

To further explore our hypothesis that maternal mortality rates are positively correlated with neonatal death rates, nutritional deficiency death rates, and protein-energy death rates, a series of 95% confidence intervals were created to examine the strength of their linkage.

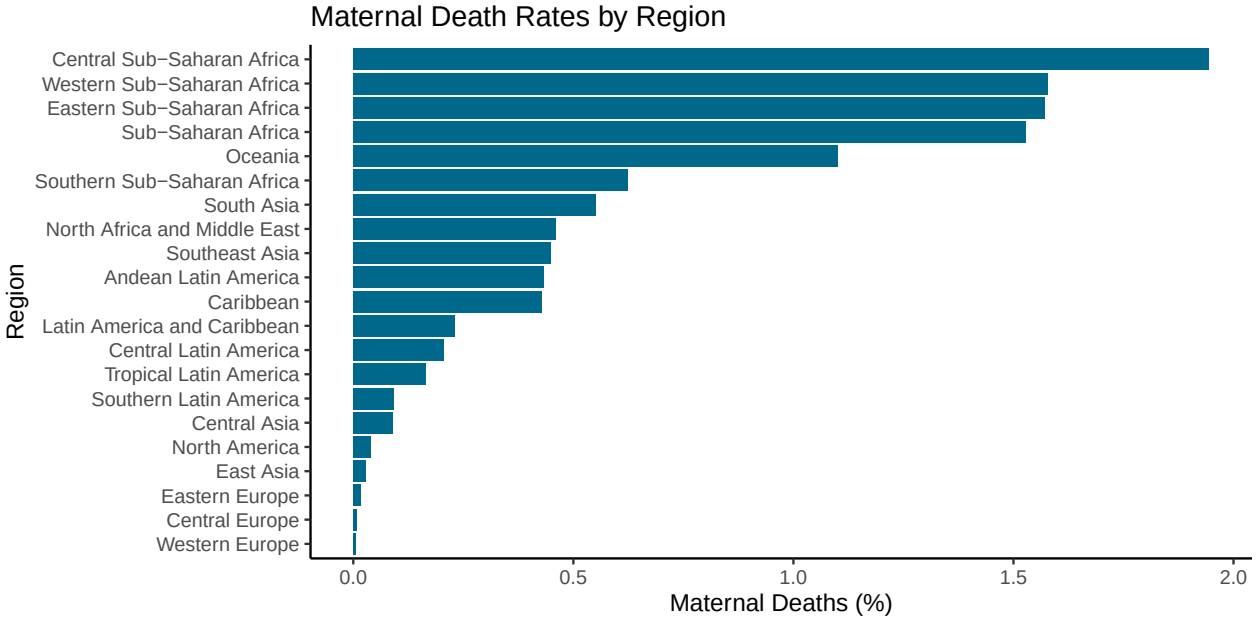
We then used linear regression to understand whether or how neonatal, nutritional, and/or protein malnutrition death rates could be used to predict maternal death rates. We also calculated the adjusted r-squared value for the first two regressions to understand how they compared to each other.

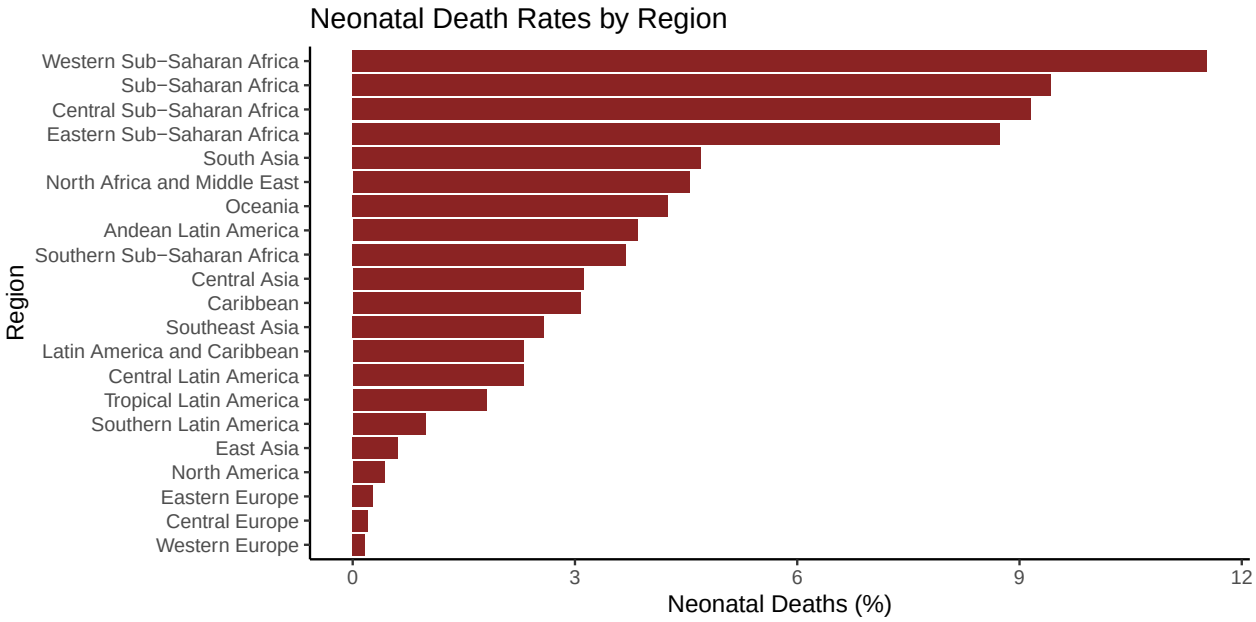
To examine the variables in a more granular manner, we turned to Sub-Saharan Africa, which our exploratory data showed to be the region with the highest maternal and neonatal death rate among all regions. We examined maternal mortality across sub-regions of the area, and then made a linear model. This model. used a new dataset we created `subsah`, which isolated the 46 countries the Lancelot classified as Sub-Saharan African (still in 2016). This model only explored the effect of deaths from nutritional deficiencies on the maternal death rates, as those numbers are tied to a causal relationship that has been posited by other sources.

RESULTS AND ANALYSIS

Exploratory Analysis of Regional Death Rates

Before we analyzed the relationships, we looked at the overall character of the 2016 data. Below are visualizations of maternal death rates and neonatal death rates by region

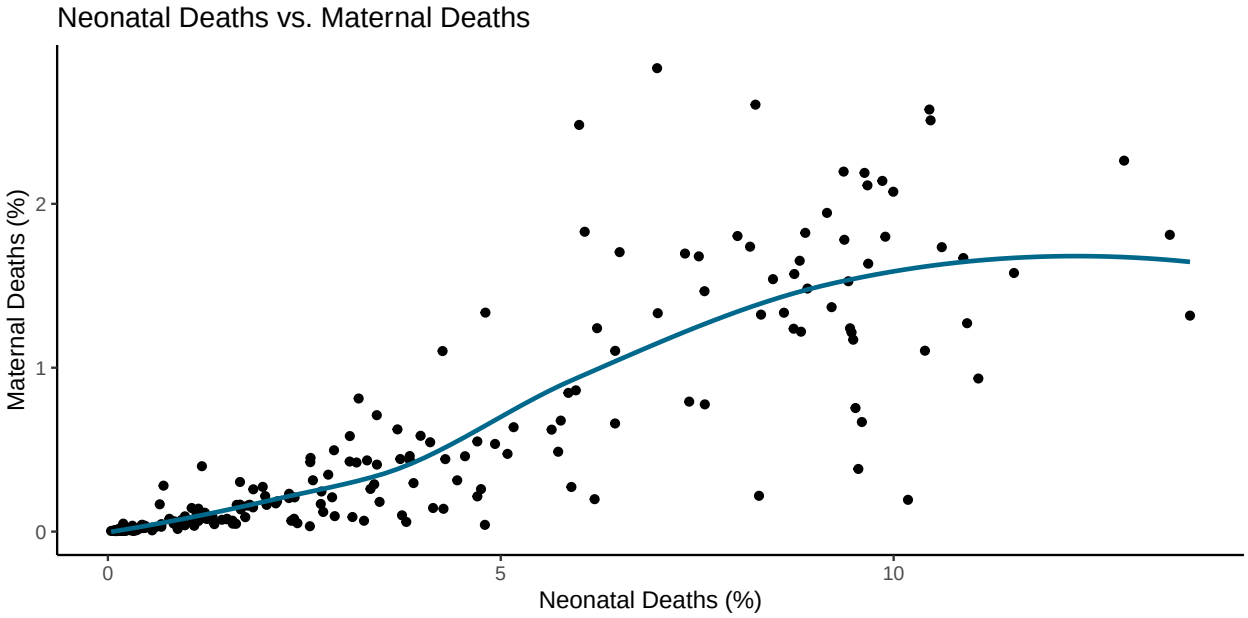




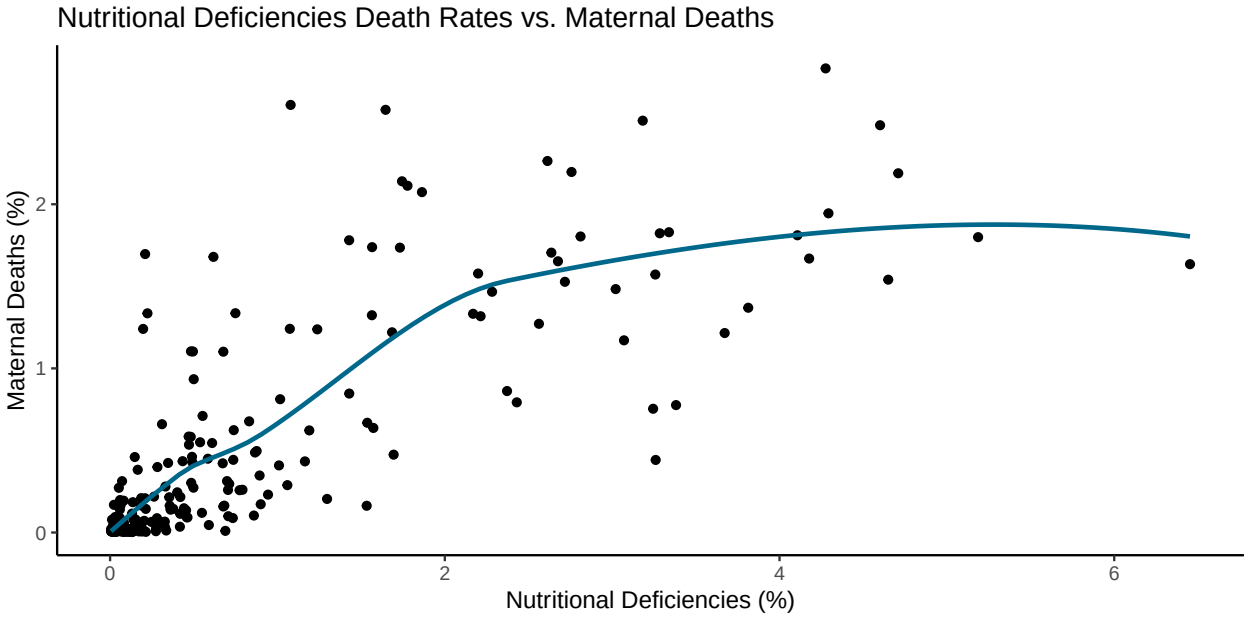
The four most prevalent bars for maternal deaths and for neonatal deaths are all various parts of Sub-Saharan Africa. The fifth highest bar for maternal deaths is also notable though - Oceania.

Death Rate Relationships

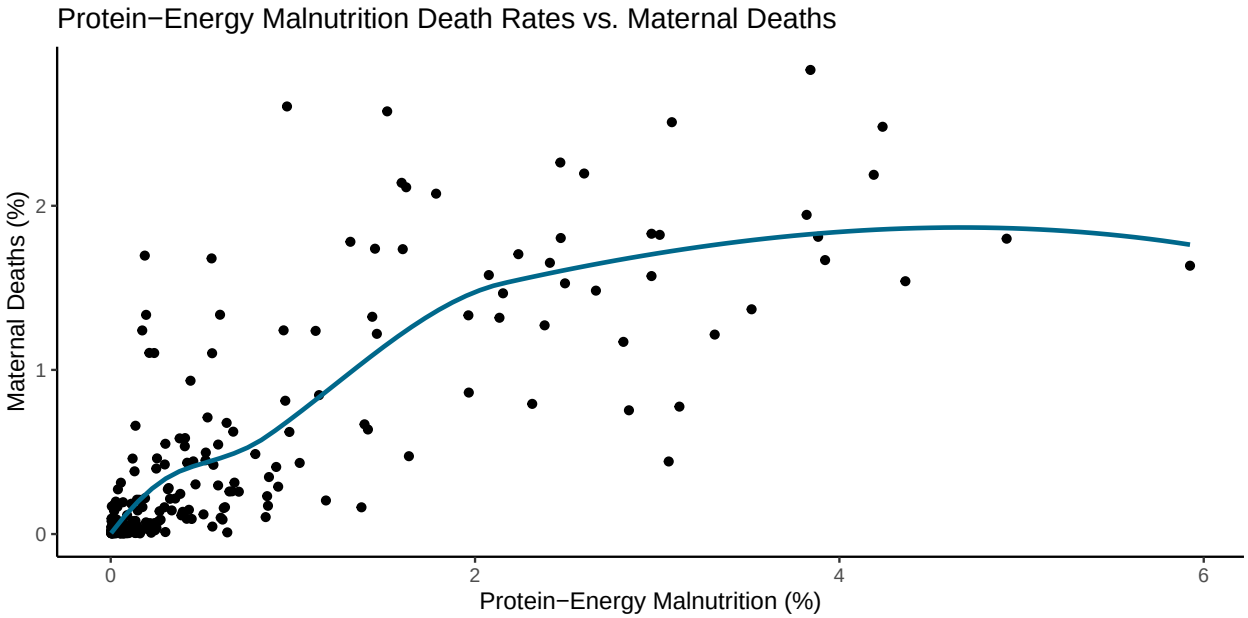
Moving forward, we examined the relationship of each of the variables, neonatal deaths, nutritional deficiencies, and protein-energy-malnutrition, to maternal deaths. The plots created in this investigation are below.



The relationship between neonatal deaths and maternal deaths appears to have a positive correlation. There are some outliers, but most places seem to be clustered at low values for both variables. The form of the curve appears to be fairly linear and has a moderately strong relationship.



The relationship between the nutritional deficiencies death rate and maternal death rate appears to have a positive correlation. It seems largely linear, but levels out as the nutritional deficiencies death rate rises. There is a relatively wide spread to it, with a cluster near zero for both values, but there are very few low maternal death rates in places with high nutritional deficiency death rates.



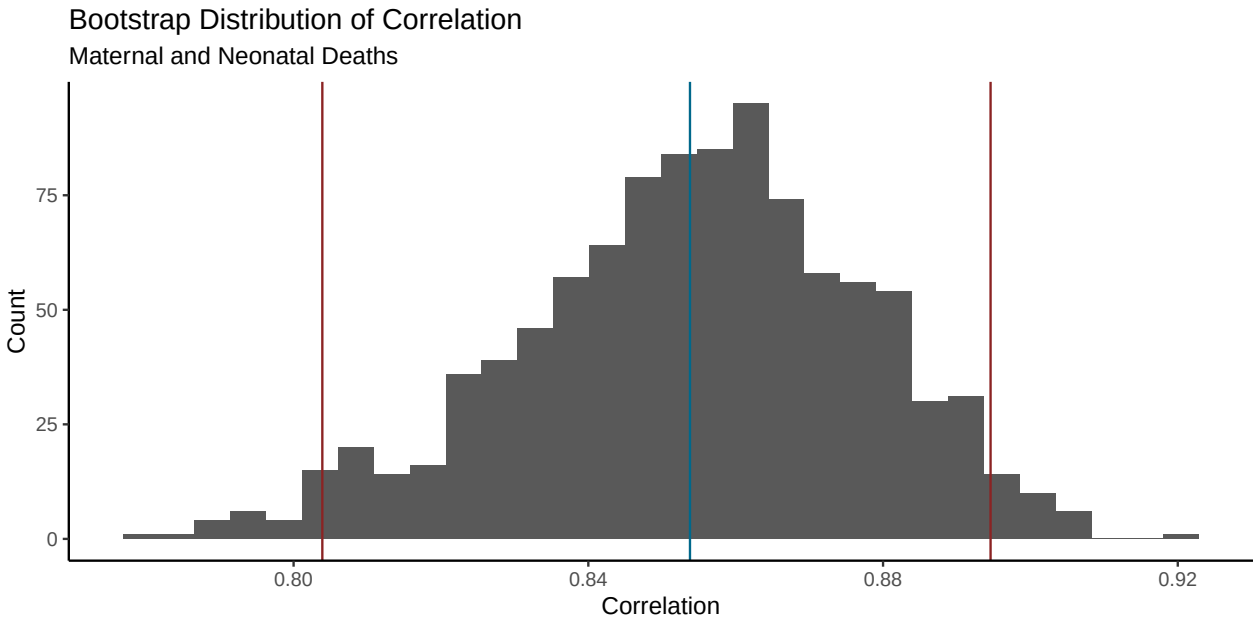
The relationship between the death rate of protein-energy malnutrition and maternal deaths appears to have a positive correlation. The shape is generally similar to the nutritional deficiencies one: a relatively wide spread, with a cluster near zero, but with very few low maternal death rates in places with high protein-energy malnutrition deficiency death rates.

Correlation Confidence Intervals

Next, in order to examine our hypothesis that maternal mortality rates are positively correlated with certain diseases and causes of deaths such as those listed above, a series of bootstrap distributions will be created to understand the relationship between the variables more specifically. We will not be performing formal hypothesis tests due to the constraints of our data. We hypothesize that maternal mortality rates will be correlated with mortality rates from nutritional deficiencies, neonatal deaths, and protein-energy malnutrition.

Table 1: Correlation between Neonatal Deaths and Maternal Deaths and 95% Confidence Interval

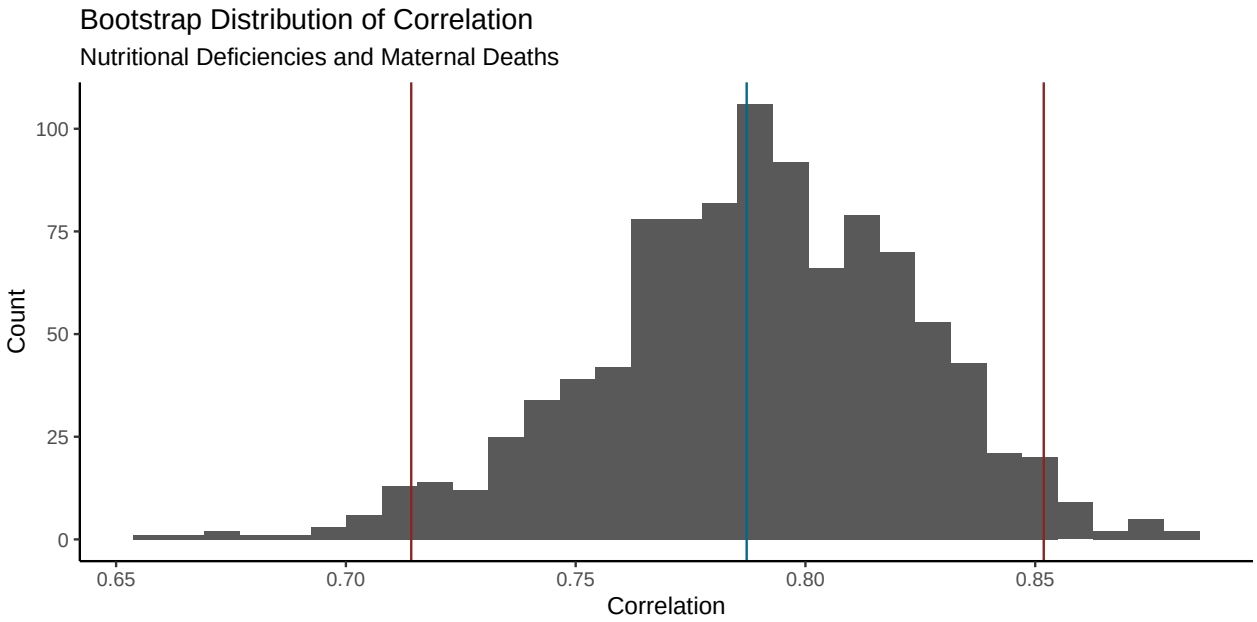
Correlation	Lower-Bound	Upper-Bound
0.854	0.804	0.895



The correlation coefficient for the relationship between neonatal death percentage and maternal death percentage is 0.854, with a 95% confidence interval of 0.804 and 0.895. This indicates that the data demonstrates a relatively strong, positive correlation between the neonatal death percentage and the maternal death percentage in 2016.

Table 2: Correlation between Deaths from Nutritional Deficiencies and Maternal Deaths and 95% Confidence Interval

Correlation	Lower-Bound	Upper-Bound
0.787	0.714	0.852

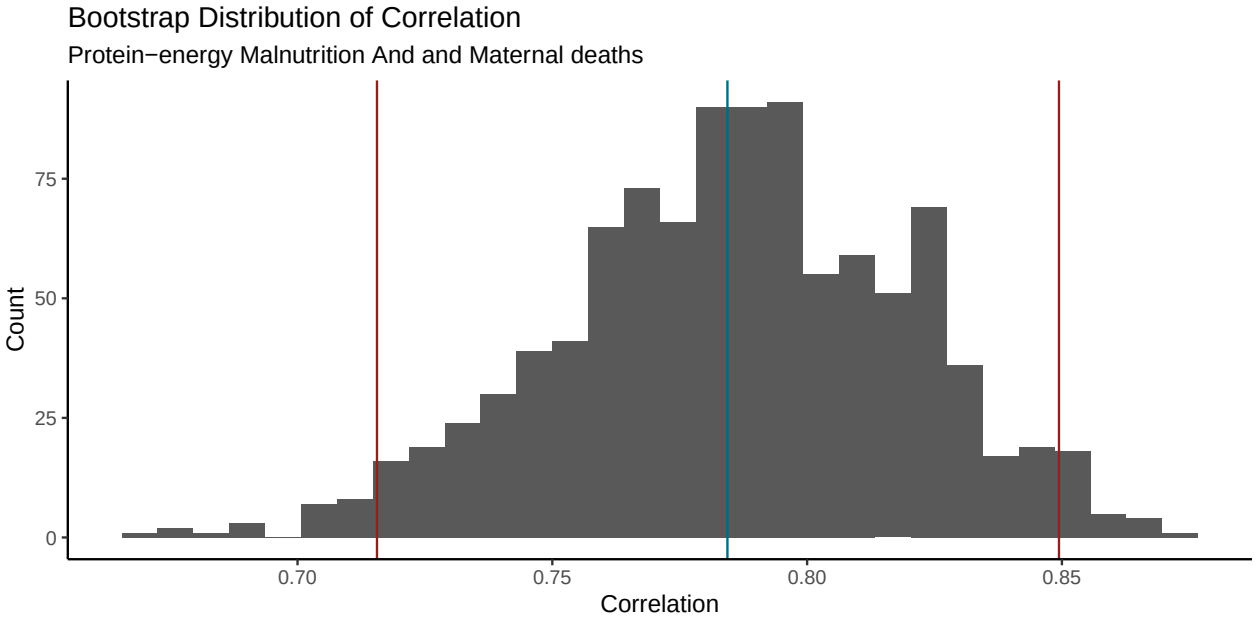


The correlation coefficient for the relationship between the nutritional deficiencies death percentage and the maternal deaths percentage in this data is 0.787, with a 95% confidence interval of 0.714 and 0.852. This

indicates that the data demonstrates a relatively strong, positive correlation between the percent of deaths from nutritional deficiencies and the percent of maternal deaths in 2016.

Table 3: Correlation between Deaths from Nutritional Deficiencies and Protein-Energy Malnutrition and 95% Confidence Interval

Correlation	Lower-Bound	Upper-Bound
0.784	0.716	0.849



The correlation coefficient of the protein-energy malnutrition death percentage and the maternal death percentage in this data is 0.784, with a 95% confidence interval of 0.716 and 0.849. This indicates that the data demonstrates a relatively strong, positive correlation between the percent of deaths from protein-energy malnutrition deficiencies and the percent of maternal deaths in 2016.

All of the 95% confidence intervals for these values overlap, indicating that they are all potentially linked to the maternal death percentage to the same degree. They are also all relatively high correlation values, pointing towards a trend where, generally, as each death percentage increases, so does maternal death percentage.

Linear Regression Modeling

To further investigate these links, we created a model for how the death percentages of nutritional deficiencies, neonatal deaths, and protein-energy malnutrition, could be used to predict the maternal death percentage. All of the conditions for this model, and the subsequent models, were checked; these graphs can be found in the appendix.

Table 4: Model of the Effect of the Rate of Neonatal Deaths and Deaths from Nutritional Deficiencies and Protein-Energy Malnutrition on the Maternal Death Rate

Term	Estimate	Standard Error	Statistic	P-Value
(Intercept)	-0.077	0.031	-2.497	0.013
`Neonatal deaths (%)`	0.114	0.009	13.290	0.000
`Nutritional deficiencies (%)`	0.408	0.337	1.212	0.227
`Protein-energy malnutrition (%)`	-0.223	0.363	-0.614	0.540

Table 5: Explanatory Power of the Model

R-Squared	Adjusted R-Squared
0.788	0.785

$\widehat{maternal} = -0.077 + 0.114 \times neonatal + 0.408 \times nutritional - 0.223 \times protein$

About 0.785 of the variability in global maternal death rates can be explained by our model using adjusted r squared. However, there are some issues with it. The intercept does not make sense to interpret, as it is negative and there cannot be a negative death percentage, even though there are places where all of the explanatory death percentages approach 0. Additionally, the p-values for the nutritional deficiencies and protein-energy malnutrition death percentages are high, indicating uncertainty in their values (especially proten-energy).

To see if it would create a better model, we removed protien-energy from our explanatory variables (chosen for having the highest p-value).

Table 6: Model of the Effect of the Rate of Neonatal Deaths and Deaths from Nutritional Deficiencies on the Maternal Death Rate

Term	Estimate	Standard Error	Statistic	P-Value
(Intercept)	-0.071	0.029	-2.428	0.016
`Neonatal deaths (%)`	0.114	0.009	13.339	0.000
`Nutritional deficiencies (%)`	0.202	0.026	7.884	0.000

Table 7: Explanatory Power of the Model

R-Squared	Adjusted R-Squared
0.788	0.786

$\widehat{maternal} = -0.071 + 0.114 \times neonatal + 0.202 \times nutritional$

This model seems to represent maternal mortality better as the p-value for nutritional deficiency is much lower here, and the adjusted r-squared is slightly higher, which is an indication of higher explanatory power despite having fewer variables. However, the intercept is still negative, and therefore does not make sense to interpret as there can not be a negative death percentage.

This model means that for everyone 1% increase in deaths from nutritional deficiencies, holding all variables constant, we expect a 0.202% increase in maternal deaths, and for every 1% increase in neonatal deaths, holding all variables constant, we expect a 0.114% increase in maternal deaths.

Returning to our preliminary data analysis, we will now investigate the relationship between nutritional deficiency deaths in the region with the highest percent of maternal deaths in 2016 - the countries in Sub-Saharan Africa. For this analysis we will concentrate only on nutritional deficiency deaths, as health organizations have previously noted a causal relationship between malnutrition and maternal mortality.

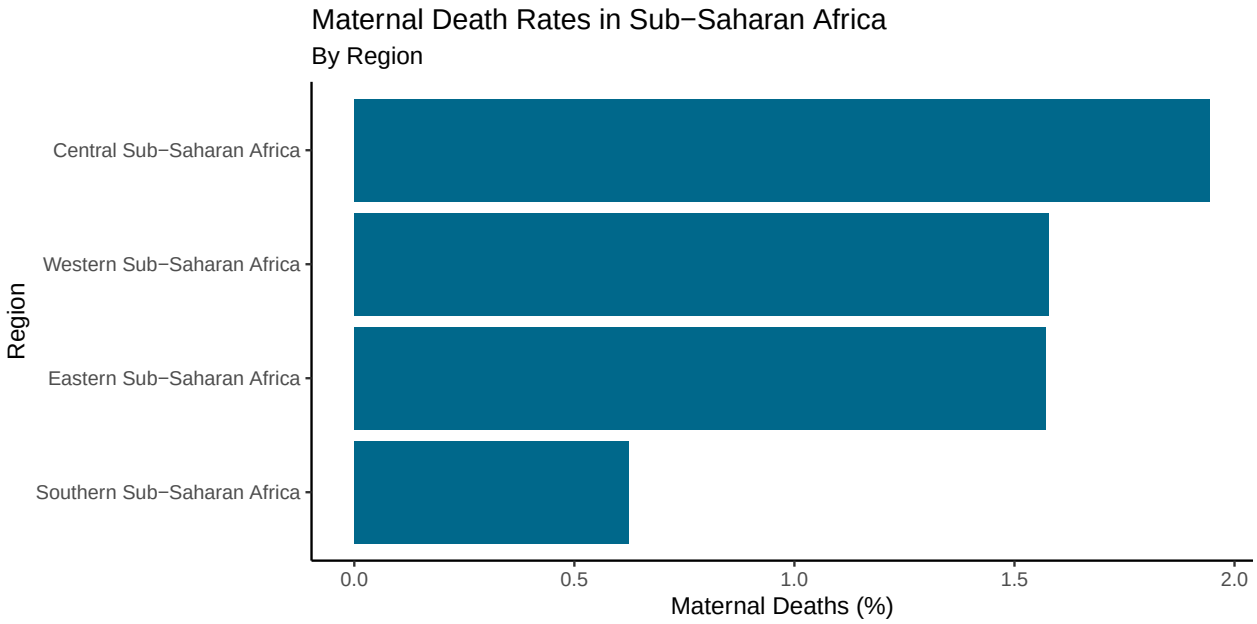


Table 8: Model of the Effect of the Rate of Deaths from Nutritional Deficiencies on the Maternal Death Rate in Sub-Saharan Africa

Term	Estimate	Standard Error	Statistic	P-Value
(Intercept)	0.920	0.190	4.829	0.000
`Nutritional deficiencies (%)`	0.221	0.066	3.353	0.002

Table 9: Explanatory Power of the Model

R-Squared
0.788

$\widehat{maternal} = .920 + .221 * nutritional$

The values in this model all have small p-values, and this intercept makes sense to interpret (which is valid to do because multiple of the nutritional deficiency death rates in this data set are relatively close to zero).

The model indicates that in a country in Sub-Saharan Africa with a 0% nutritional deficiencies mortality, .920% of deaths would be maternal deaths. This implies that there are other factors in Sub-Saharan Africa that are related to the higher mortality rates than just deaths from nutritional deficiencies, although the r-squared value of 0.788, 0.786 does mean that about 78.764% of the variability in the maternal death rates can be explained by the nutritional deficiencies death rate. Additionally, the model includes a relationship between the two where, for every 1% increase in deaths from nutritional deficiencies, there is an increase of .221 in the predicted maternal death percent. However, these things could also be because of other variables affecting both of the values.

DISCUSSION AND CONCLUSION

This analysis and the data set must be understood through the lens of some of its limitations. The first weakness of the dataset as a whole is that many of the countries with recorded data may not have ideal reporting conditions for keeping track of deaths throughout the country. For example, a specific region or country may not have adequate funds or infrastructure to collect data on all of its population. However, to accommodate for this, The Lancet used a variety of metrics to adjust for these discrepancies, which includes using a reliability rating that depends on an institutions ability to collect accurate data. Therefore, not all of the death rates may be entirely accurate for a country or region. Furthermore, all of the data is expressed as percents - we don't have access to the raw data, and therefore we cannot analyze things like the concrete number of people that died in each way. A country/region with a smaller population may have a higher death percent than a country/region with a larger population, even though more people died of a certain disease or issue in the country/region with a larger population. Another big issue with the data is that the values are not necessarily independent of each other. It is unclear how the dataset handles multiple causes of death. If it does account for this, it is not guaranteed that this would be accurately reported to The Lancet. Finally, our analysis only looked at the most recent year we had access to (2016) and we do not know how reflective that year was of general trends in other years.

In terms of our analysis, one weakness was the wide confidence intervals we had for the correlation values. For example, the correlation confidence interval of Nutritional Deficiencies and Maternal Deaths had a lower bound of 0.714 and an upper bound of 0.852; having a correlation value of the former indicates a notably lower degree of linkage than the latter, so not knowing this more precisely restricts our understanding. Furthermore, we do not know why these relationships exist (what is the mechanism behind their correlation or causation). Finally, in the conditions to check for our multiple linear regression, none of the graphs were ideal in terms of fulfilling the condition. However, they were generally adequate enough to be comfortable with linearity, equal variance, independence, and normality.

Despite these limitations, we can draw some conclusions from the data. For one, it appears that, for the data we analyzed, there is a clear positive relationship between maternal deaths and both neonatal deaths and deaths from nutritional deficiencies, as well as a possible one between maternal deaths and protein-energy malnutrition deaths. One potential explanation for the negative relationship with protein-energy in the first model (despite its initial positive correlation) is that that model also accounted for deaths from general nutritional deficiencies. Protein-energy may not be the most damaging form of malnutrition, so when it was high in the place of others, maternal mortality may have decreased. There was also a high amount of uncertainty in this number though. We can also see that this relationship between deaths from nutritional deficiency and maternal deaths holds in Sub-Saharan Africa where the latter is a prominent issue. This indicates that interventions that address problems at the root of both causes of death could have a powerful impact. These findings also higlight the importance of broadly conceptualizing health interventions. It is not just emergency or clinical care that can improve people's wellness - it is also about having sustainable, rigorous supply chains to maintain communities and increase ease of access to necessities

If we were to do this project again, we think it would be interesting to look at this data over a range of time, rather than just during one year. It is also an incredibly rich dataset, so there are many other potential things we could examine within the context of maternal mortality, or between each other. Being able to do more outside research could have also been fruitful in that we could have better understood the relationships we were uncovering. Specifically, examining other variables that have a known relationship to babies and female bodies could have been interesting, like diarrheal diseases for the former and HIV/AIDS for the latter. One interesting route could be to look at whether there are issues that affect maternal mortality but not neonatal, and vice versa.

REFERENCES

https://www.thelancet.com/gbd/about

https://ourworldindata.org/what-does-the-world-die-from

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30925-9/fulltext

https://www.ncbi.nlm.nih.gov/books/NBK201704/

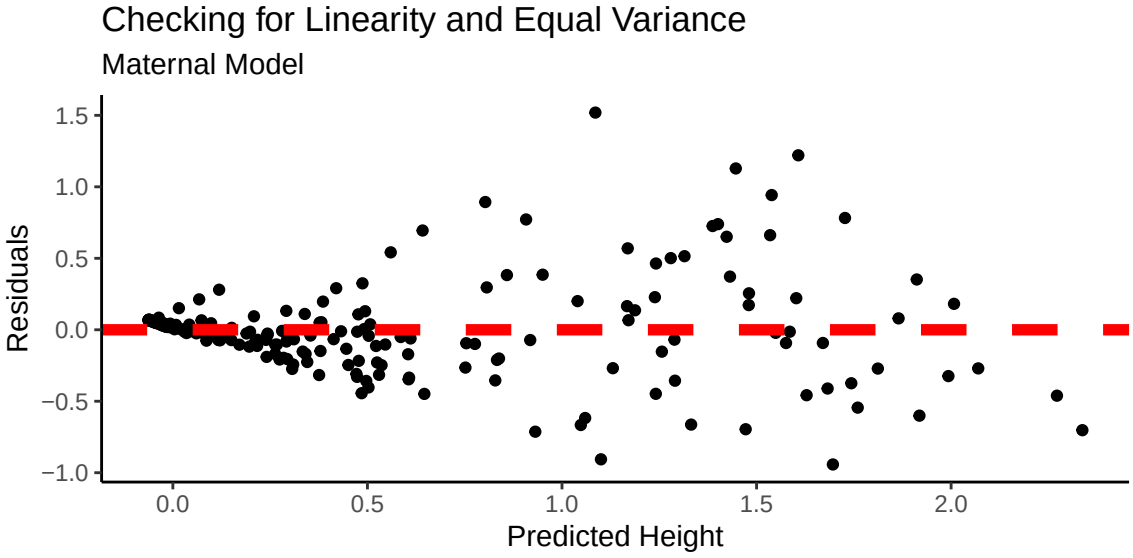
https://www.un.org/sustainabledevelopment/sustainable-development-goals/

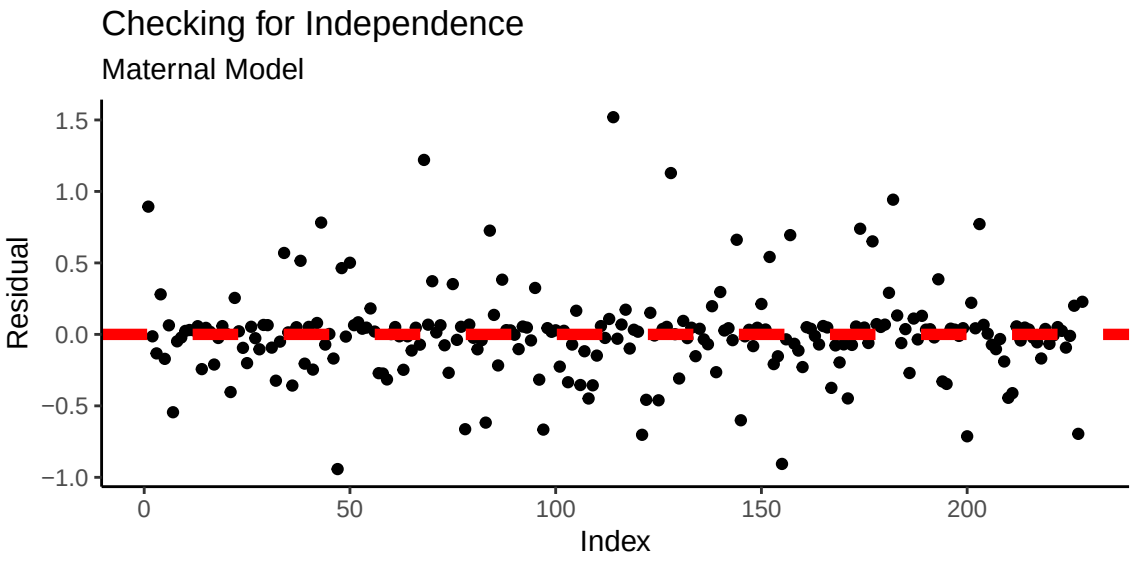
https://www.thelancet.com/cms/10.1016/S2214-109X(19)30374-2/attachment/6a9c1d93-e0a1-4b7d-ba2d-f10c67026006/mmc1.pdf

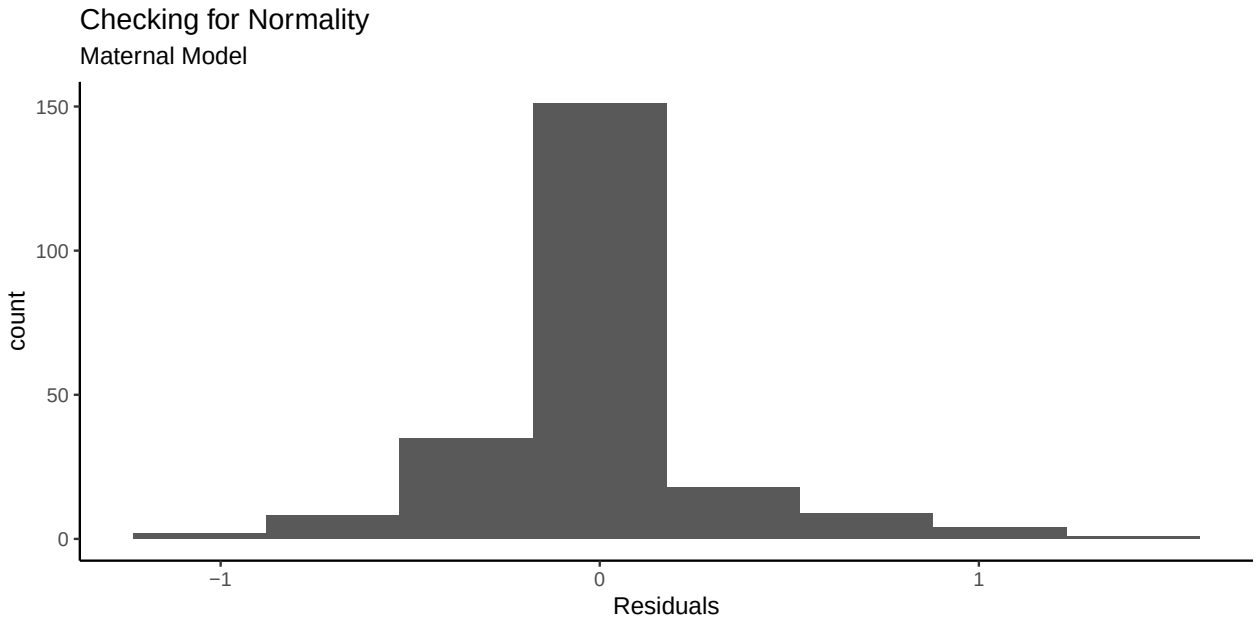
https://www.usaid.gov/sites/default/files/documents/1864/role-of-nutrition-preventing-child-maternal-deaths.pdf

APPENDIX

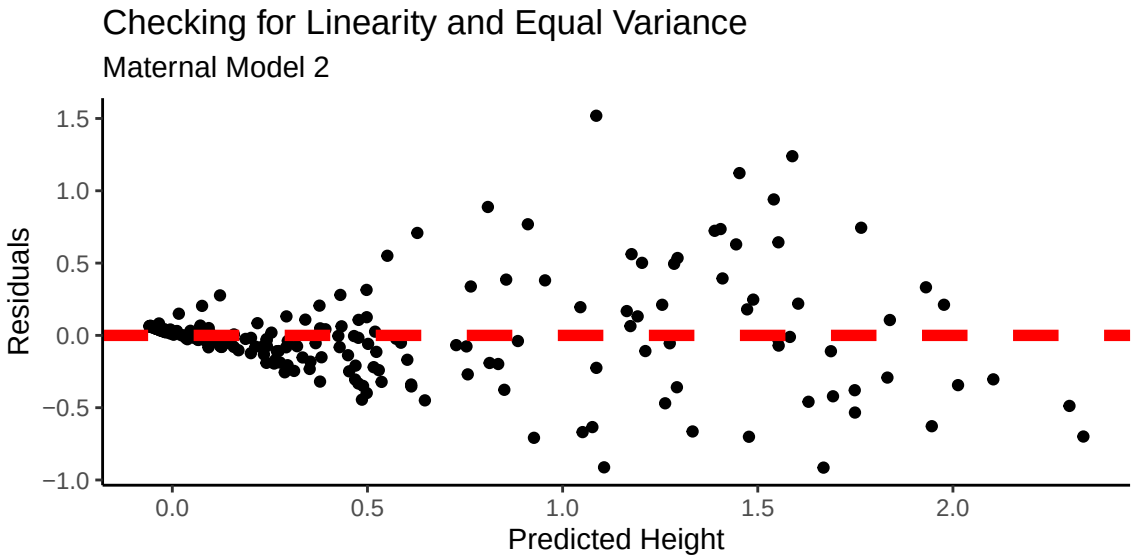
Checking Conditions for Maternal Model

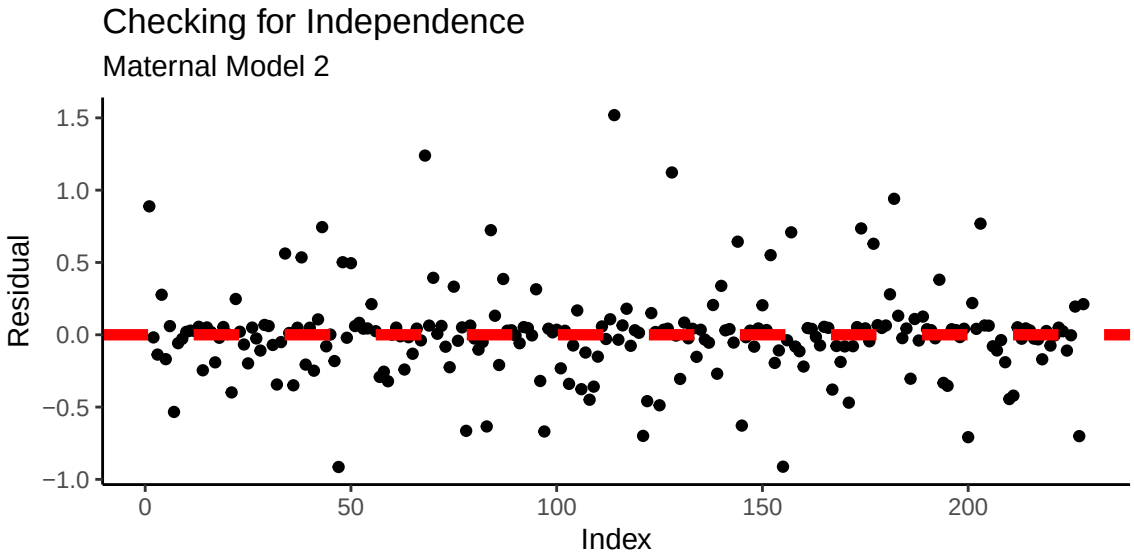


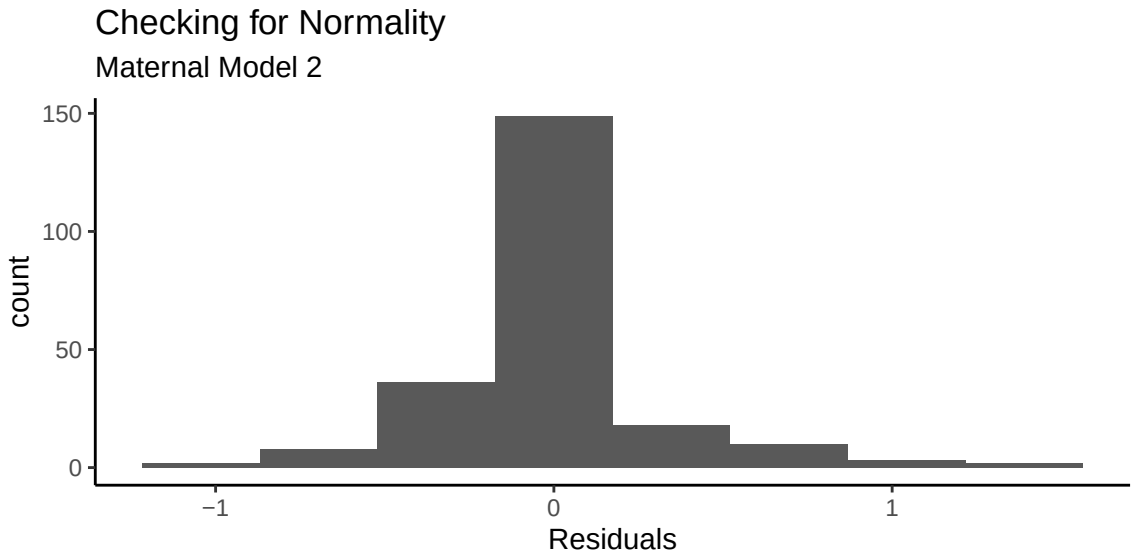




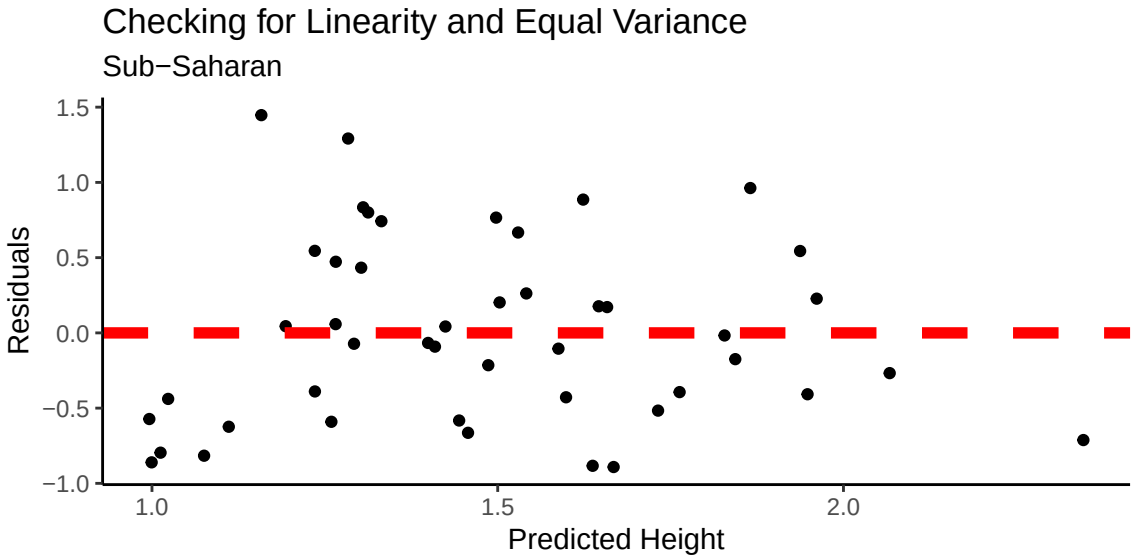
Checking Conditions for Maternal Model 2

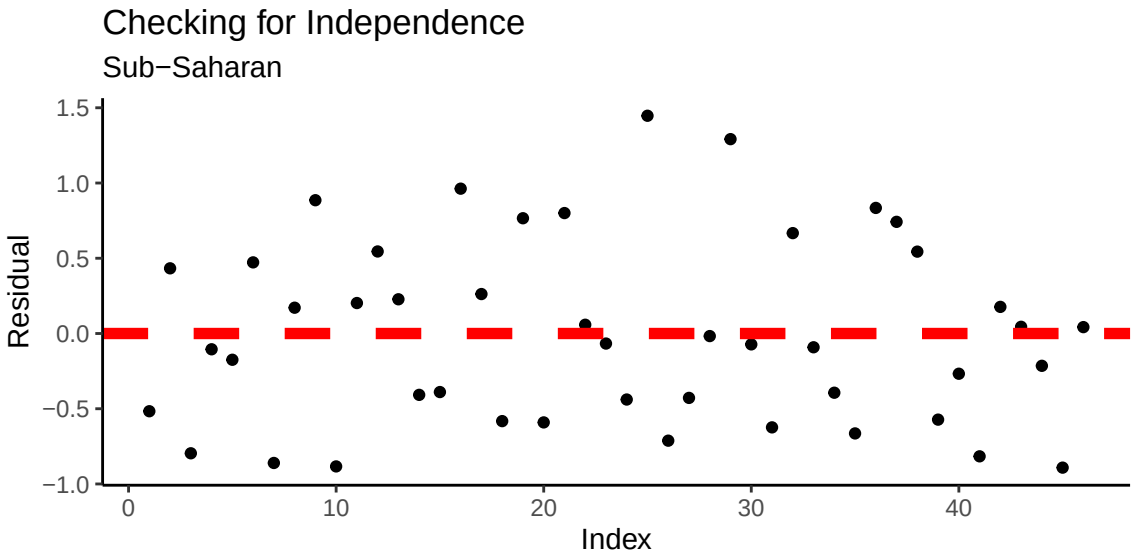


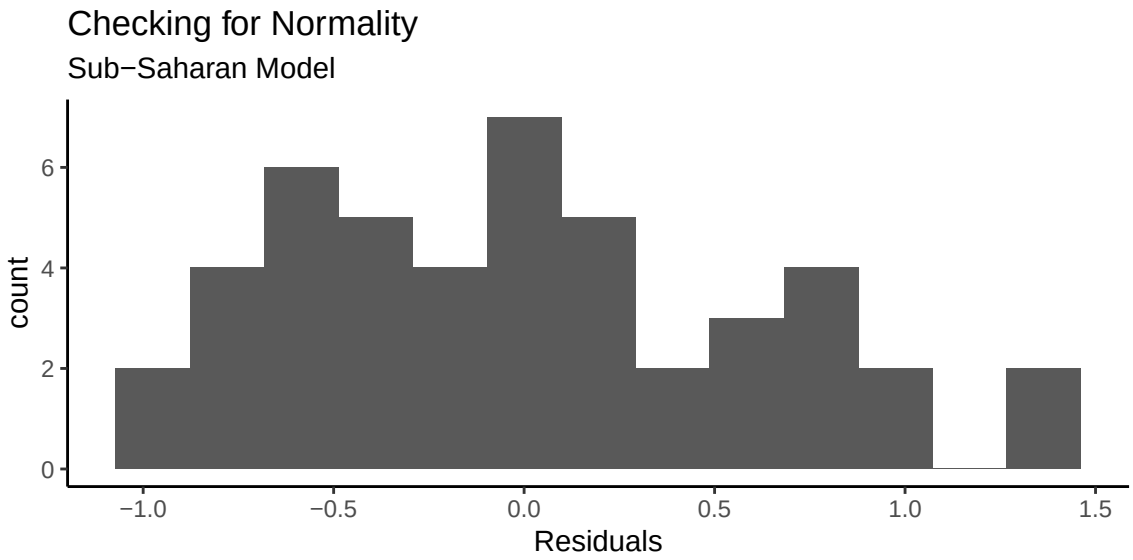

Checking for Independence
Maternal Model 2
Residual
1.5
1.0
0.5
0.0
−0.5
−1.0
0
50
100
150
200
Index


Checking for Normality
Maternal Model 2
count
150
100
50
0
−1
0
1
Residuals

Checking Conditions for Sub-Saharan Model (sub_model)






Checking for Normality
Sub−Saharan Model
count
0
2
4
6
−1.0
−0.5
0.0
0.5
1.0
1.5
Residuals