INFANT & MATERNAL MORTALITY:

Natal Health Predictors in the CIA Factbook

G. Adolph
S. Blanco
W. Thoss

Alex J. Quijano
MATH 141: Introduction to Probability & Statistics
15 November 2021
Reed College

Introduction

The CIA factbook contains data from every country around the world, with important statistics such as area, population, internet users, migration rate, among others. In this study we are interested in the statistics for the maternal mortality rate and the infant mortality rate. Specifically, we wanted to see if there is any kind of relationship between the two. To determine this, we performed both a single linear and multiple linear hypothesis test to see the correlation between the variables which would either affirm or contradict our alternative hypothesis. We hypothesized that the infant mortality rate is a positive predictor of maternal mortality rate in a given country and thus that countries with higher infant mortality rates tend to have higher maternal mortality rates. We also hypothesized that maternal mortality rate would be the most significant predictor variable of infant mortality rate.

Data Exploration

Research Question

Given that infant and maternal mortality rates are commonly used indicators of maternal health services in a country, we investigate the following questions (*Infant Mortality | Maternal and Infant Health | Reproductive Health | CDC*, 2021):

1. Is there a positive linear relationship between a country's infant mortality rate and the maternal mortality rate?
2. Is a country's maternal mortality rate the most significant predictor of the infant mortality rate in a multiple regression model?

Hypotheses

Null Hypothesis$_1$ $[H_{0\_1} : b_1 = 0]$: There is no correlation between the infant mortality rate of a country and its maternal mortality rate.

Alternative Hypothesis$_1$ $[H_{A\_1} : b_1 > 0]$: The infant mortality rate will have a positive effect on the maternal mortality rate.

Null Hypothesis$_2$ $[H_{0\_2} : b_{maternal\_mortality\_rate} = b_{population} = b_{population\_growth\_rate} = b_{net\_migration\_rate} = b_{internet\_users}]$: The effect maternal mortality rate has on infant mortality rate is equal to the effect that population, population growth rate, net migration rate, and the number of internet users has on infant mortality rate.

Alternative Hypothesis$_2$ *[*H_{A_2} *:* $b_{maternal\_mortality\_rate}$ *>* *[*$b_{population}$ *AND* $b_{population\_growth\_rate}$ *AND* $b_{net\_migration\_rate}$ *AND* $b_{internet\_users}$*]*: The maternal mortality rate will have a greater effect on infant mortality rate than population, population growth rate, net migration rate, and the number of internet users.

Methods

A simple linear regression was performed to examine the isolated relationship between infant and maternal mortality rates without considering the effect of any other variables (Fig. 1). The infant mortality rate was observed to be a statistically significant predictor of maternal mortality rate given the simple linear models produced (Eq. 1) (Adjusted $R^2 = 0.6189$, $F = 296.5$, $df = 1, 181$, $p < 2.2x10^{16}$). The R^2 value suggests that only about 62% of the variability in the data is represented by the single regression model. The variance in the residuals tends to be greater for larger values thus the residuals exhibit slight heteroscedasticity which suggests that the greater the value, the more varied it is from the standard distribution (Fig. 2). Because the distribution of residuals is not severely different from the standard normal distribution, we did not perform any normalizing transformations on the data (Fig. 3). If the data in Figure 2 & Figure 3 were not as densely organized around the reference lines, we would not be able to conduct statistical analyses without performing normalizations to correct the varied amount of variance throughout the data.

Equation 1.

$$Infant\ Mortality\ Rate\ =\ [14.437\ +\ (7.5099\ \times\ 10^{-2})\ *\ Maternal\ Mortality\ Rate]$$

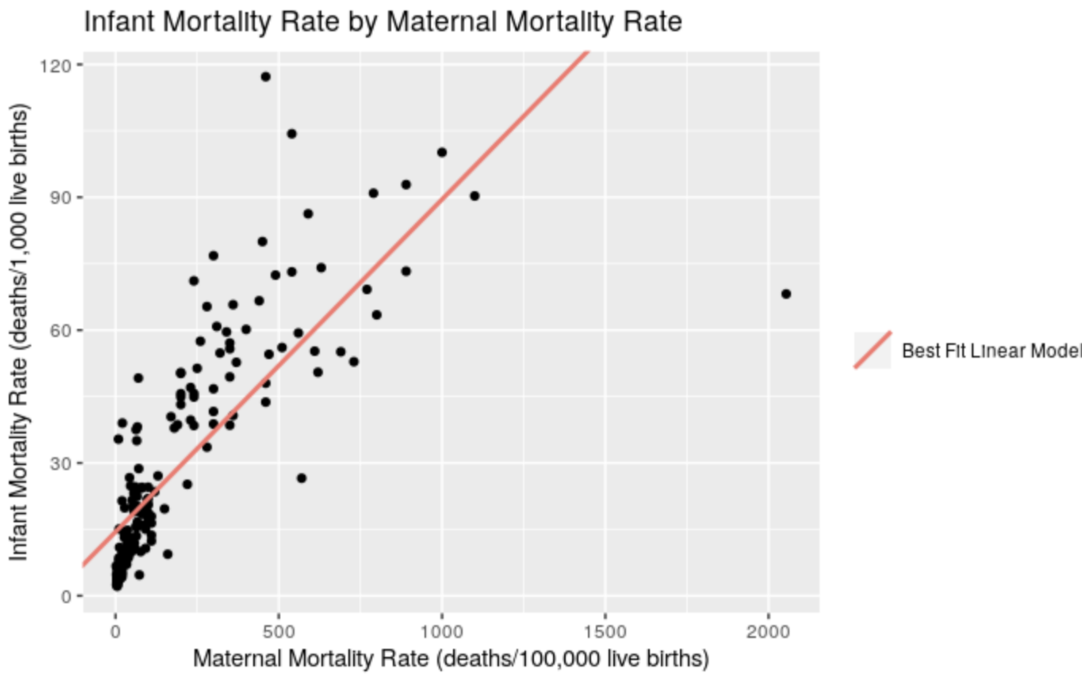

Figure 1. Scatterplot of infant mortality rate (deaths per 1,000 live births) by maternal mortality rate (deaths per 100,000 live births) with the best fit linear model (Eq. 1)

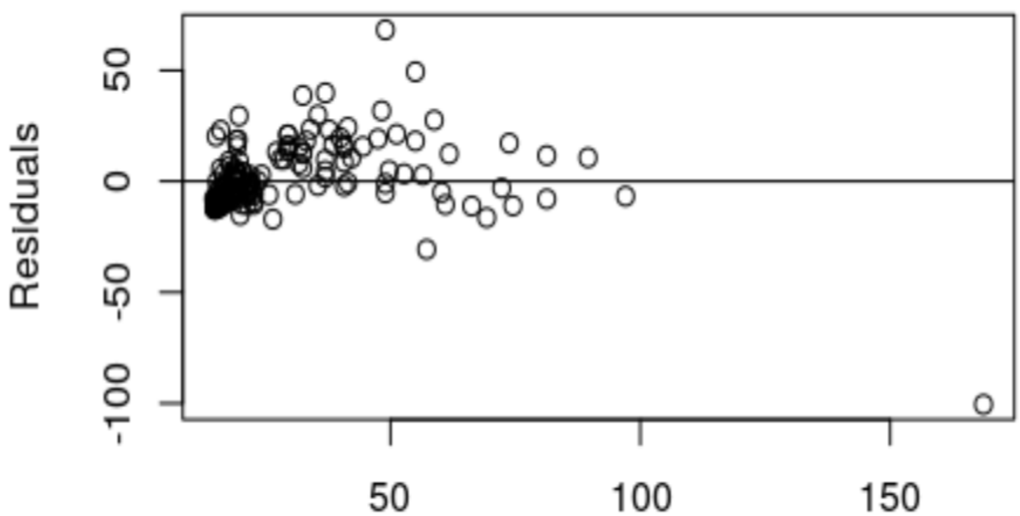

Figure 2. Scatterplot of the residuals for the single linear model (Eq. 1)

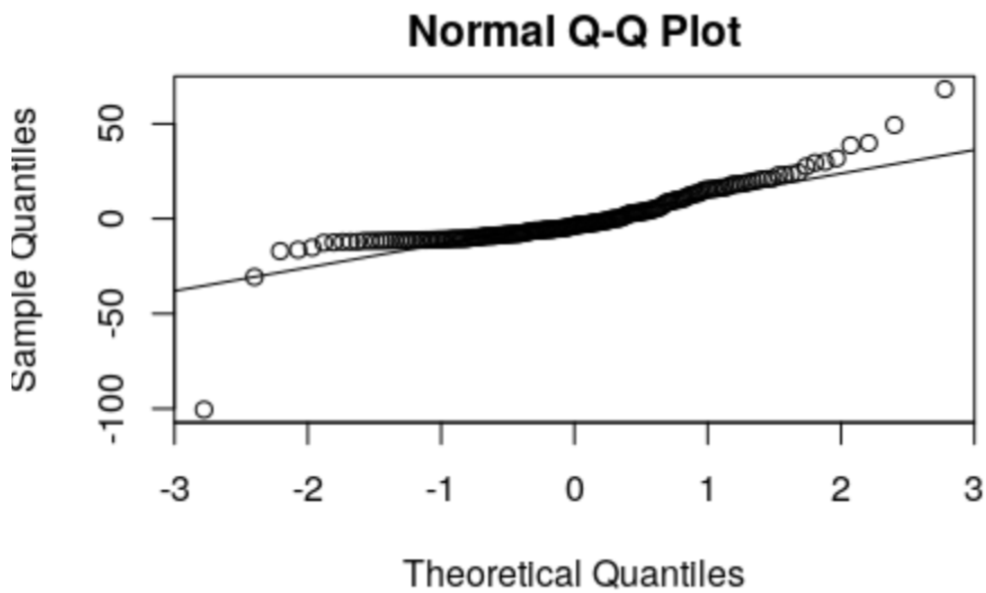

Figure 3. Normal quantile-quantile plot for the single linear model (Eq. 1)

A multiple linear regression was performed to determine the impact other variables might have on infant mortality rate and the significance of maternal mortality rate as a predictor. A multiple linear regression was also performed to observe whether the predictors of infant mortality rate were similar to those of maternal mortality rate. Infant mortality rate ($p < 2x10^{-16}$) was the only observed statistically significant predictor variable for maternal mortality rate (Adjusted $R^2 = 0.730$, $F = 96$, $df = 5, 171$, $p < 2.2x10^{-16}$) whereas maternal mortality rate ($p < 2x10^{-16}$), population ($p = 0.049$), population growth rate ($p = 7.19x10^{-11}$), and net migration rate ($p = 7.63x10^{-10}$) were all observed to be statistically significant predictors of infant mortality rate in the multiple regression model (Adjusted $R^2 = 0.795$, $F = 137.9$, $df = 5, 171$, $p < 2.2x10^{-16}$) (Fig. 2). The R^2 value suggests that nearly 80% of the variability in the data is represented by our multiple regression analysis. The multiple linear model produced an equation that predicts infant mortality rate as a function of each of the statistically significant predictor variables, each with its own coefficients (Eq. 2). Analysis of collinearity statistics show that there is no multicollinearity in the multiple regression model given the variance inflation factors (VIF, max = 4.024, min = 1.761, mean = 2.827, median = 2.500). The Durbin-Watson statistic (Durbin-Watson = 1.937) proves that the values of the residuals are mostly independent with only minor positive autocorrelation. The scatterplot and quantile-quantile plot of residuals, however, exhibit heteroscedastic distributions that suggest there is more variance in the larger values and that the variability is not equal throughout the dataset (Fig. 4; Fig. 5). The kernel density estimate also displays a slight right skew in the residuals, supporting the observation that variability tends to increase with the size of the data points (Fig. 6).

Equation 2.

$$Infant\ Mortality\ Rate = \{5.372 + [(7.062 \times 10^{-2}) * Maternal\ Mortality\ Rate] + [(1.818 \times 10^{-8}) * Population] + [(7.884) * Population\ Growth\ Rate] + [(-8.502 \times 10^{-1}) * Net\ Migration\ Rate]\}$$

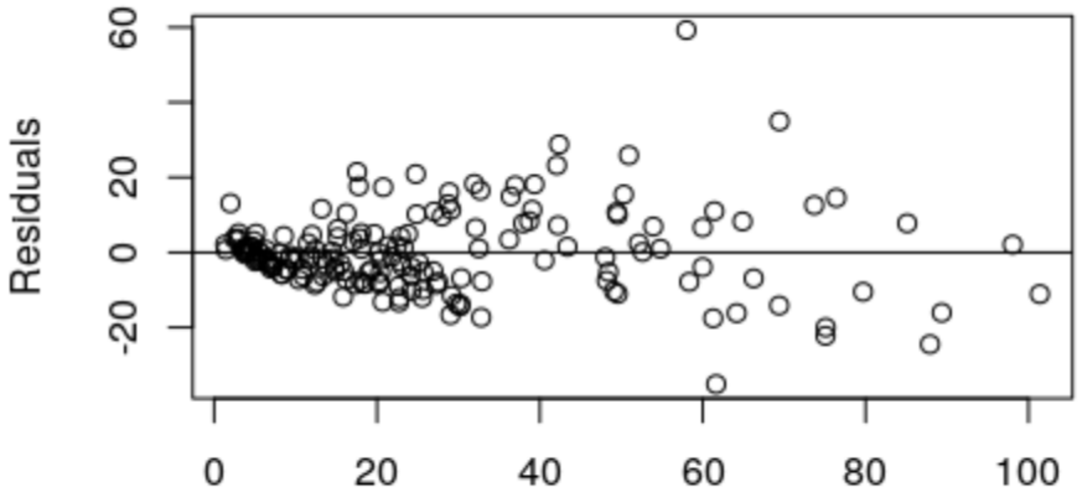

Figure 4. Scatterplot of the residuals for the multiple linear model (Eq. 2)

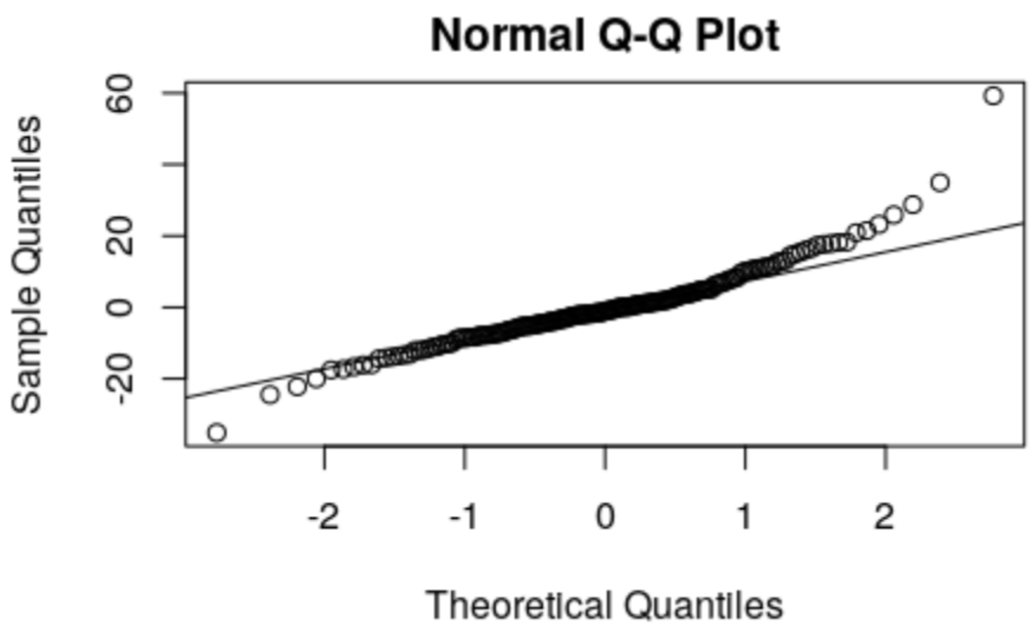

Figure 5. Normal quantile-quantile plot for the multiple linear model (Eq. 2)

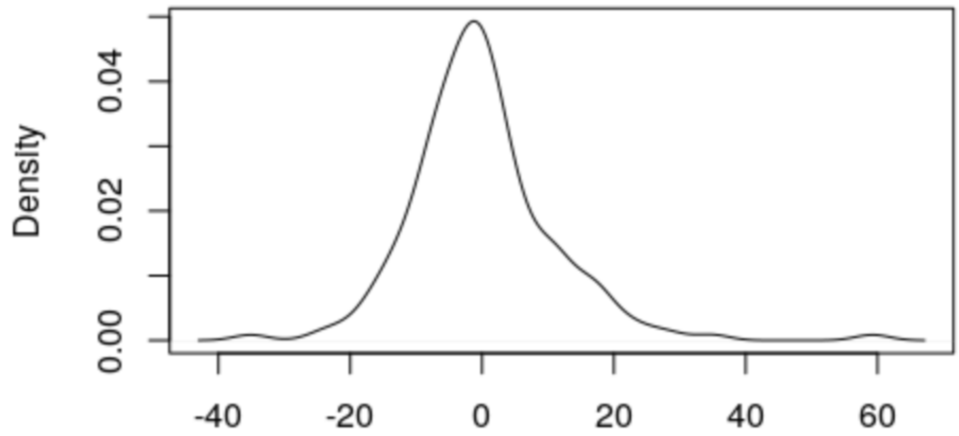

Figure 6. Kernel density estimation (N = 177, bandwidth = 2.631) of the residuals in the multiple regression model (Eq. 2)

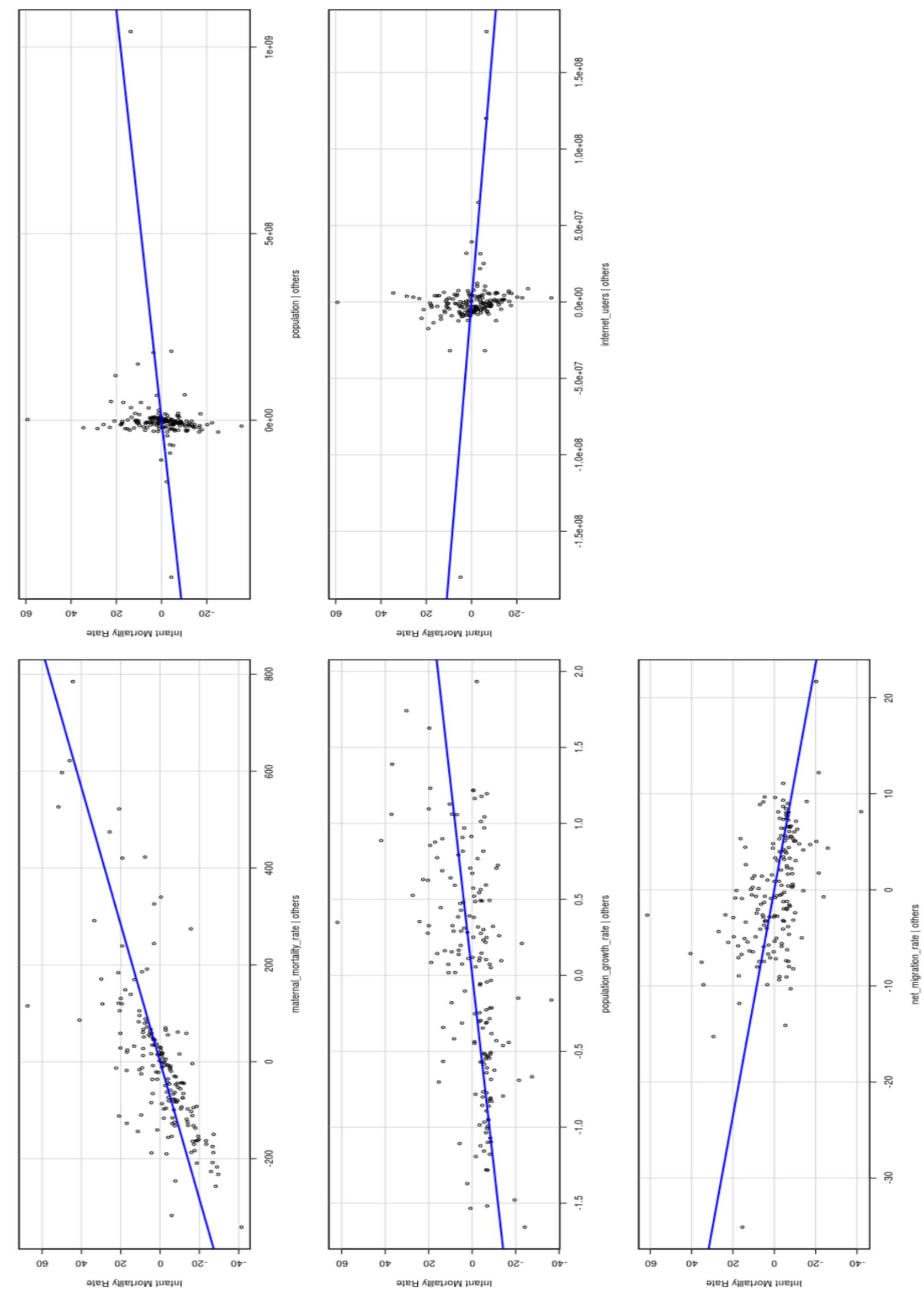

Figure 7. Partial regression plots representing the transformations of each numerical predictor variable and infant mortality rate (deaths per 1,000 live births)

Results & Discussion

More Exploration:

Randomization: For our randomization test, we will use a "success-failure" model, depending on the value of maternal and infant mortality rates. We will choose a point and call anything above that value a "success" and anything below a "failure" so as to have two categorical variables by which to complete the shuffling.

We are using proportions to examine the data. Our point estimates are the true proportions of infant mortality rate to maternal mortality rate. Our initial findings are that there seems to be a positive correlation between infant mortality and maternal mortality rates. As the rate of infant mortality increases in each country, there is a higher likelihood of the maternal mortality rate increasing. Our point estimates are the maternal mortality rate and the infant mortality rate of each country we are looking at.

Based on the data provided, we can see that the relationship between maternal mortality and infant mortality has a positive correlation. This signifies that when a country has a higher maternal mortality rate, it will also have a higher infant mortality rate. A positive correlation between the two variables makes sense because poor medical care for mothers would likely translate into poor medical care for the children. Several other variables that could be explored include the country of all of the subjects we are looking at. Then, within each country, we could look at the population and population growth rates. After some initial research, there doesn't seem to be any relationship between population and infant or maternal mortality rates. The distribution appears to be quite random. There are also several variables that aren't provided by the CIA factbook that would still be useful to look into. The GDP per person of a nation would give us a good sense of whether wealth plays a role in natal health. It could also be interesting to break the countries down by geographical location, and see if any patterns arise from such an organization.

Conclusions

Following numerous statistical tests upon the data, we can conclude that there is a positive linear relationship between the maternal mortality rate and the infant mortality rate of a country. This disproves our null hypothesis and supports our alternative hypothesis that there is a linkage between the two variables. All of the tests that we performed supported the alternative hypothesis, especially the adjusted R^2 value of 0.795, an exceptionally high value, indicating that 79.5% of the variation in infant mortality rate correlates with changes in the maternal mortality rate of a country. Ultimately this makes sense, as maternal health and infant health are undoubtly linked, especially in birth. It's also true that nations that lack medical care would struggle with both maternal and infant health. The more interesting possible conclusion would have been that there is no correlation between the two variables. Through further tests, we also concluded that there were no significant linear relationships between infant mortality rate and any of the other variables presented by the CIA factbook.

Appendix

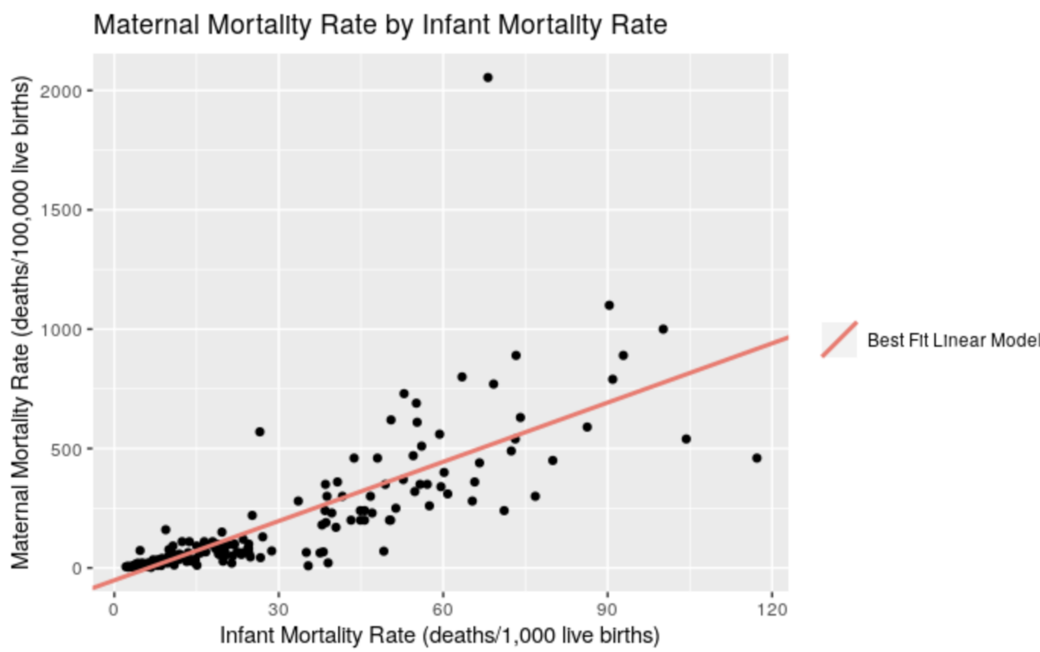

Figure 8. Scatterplot of maternal mortality rate (deaths per 100,000 live births) by infant mortality rate (deaths per 1,000 live births) with the best fit linear model (Eq. 3) (inverse axes of

Equation 3.

$$Maternal\ Mortality\ Rate\ =\ [-\ 51.561\ +\ 8.269\ *\ Infant\ Mortality\ Rate]$$

References

Infant Mortality | Maternal and Infant Health | Reproductive Health | CDC. (2021, September 8). https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm

LOUDON, I. R. V. I. N. E. (1991). On maternal and infant mortality 1900-1960. *Social History of Medicine*, *4*(1), 29–73.

https://academic.oup.com/shm/article/4/1/29/1651347?login=true https://www.jstor.org/stable/25767377?Search=yes&resultItemClick=true&searchText=maternal+and+infant+mortality&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dmaternal%2Band%2Binfant%2Bmortality%26so%3Drel&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3Ab756c6ec54a5d7108d59ff2672e201b5

https://www.jstor.org/stable/29789108?Search=yes&resultItemClick=true&searchText=predictors+of+infant+mortality&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dpredictors%2Bof%2Binfant%2Bmortality%26filter%3D%26so%3Drel&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3A09ede7d71e97df3c09cb605e3289db88&seq=1#metadata_info_tab_contents