Introduction

The goal of the study is to observe if there is a linear relationship between the maternal mortality rate and several other factors. The maternal mortality death is when mothers die from giving birth or other childbearing complications (UNICEF DATA, 2021). In the study reviewed under the National Center for Biotechnology Information, factors commonly associated with maternal mortality are “obstructed labor, malnutrition, poverty, overwork, lack of primary health care, parasitic disease” (Stokoe). Not only are these medical factors affecting the rate, culture factors such as education on pregnancy or sexism also play important factors in this.

As the article of a statistical study over the maternal mortality rate between 1990 to 2006 concluded, the study of maternal mortality rate is important because it is an indicator of development, culturally, economically, and medically. Countries with higher rates of maternal mortality are usually less developed countries or countries with fewer skilled professionals in this area of the medical field.

The dataset is from the International Conference on Population and Development. The dataset we chose does not contain many variables but we are still only selecting some to conduct the experiment on. We will be choosing factors that are relevant to the above mentioned factors in Introduction because they are backed by statistical analyses. It will give us a clearer idea of how the data should look like but if the data doesn’t turn out the same, we can analyze possible errors, if any.

Methods

We will be selecting Births attended by skilled health personnel, percent, Number of new HIV infections (all ages, per 1,000), Contraceptive prevalence rate (women aged 15-49, any method), Contraceptive prevalence rate, women aged 15-49, modern method, Unmet need for family planning (women aged 15-49), and Laws and regulations that guarantee access to sexual and reproductive health care, information and education (percent).

HIV infection factor is selected because pregnant women have weaker immune systems and people often pass away from the infection. Related contraceptive prevalence factors are picked because contraception prevents pregnancy. The prevention is not 100% but if a woman got pregnant with contraceptive, the baby is less likely to have problems.

Laws and regulations of reproductive health factors are picked because cultural factors also affect the maternal mortality rate. If girls and boys aren’t educated about pregnancy and childbirth, they would not have the knowledge to be prepared to face these problems. Thus increasing the chance of mortality.

We didn’t pick the lower and upper range of the maternal mortality rate because the lower and upper range of rate wouldn’t affect the maternal mortality rate. The range is only the indicator of the possible maternal mortality rate of each country. The factor Proportion of demand satisfied with modern methods, women aged 15-49 wasn’t selected because it is repetitive of two factors that we will be performing the experiment on.

Conduct a basic exploratory analysis to graph each predictor and response variable on scatter plots and observe patterns. Note if there are any outliers or skewed patterns.

Check for model assumption violations by graphing residual plots of six predictors against the response variable (maternal mortality rate) and look for any systematic patterns, clusters, fanning patterns. If there are clustered points, perform a box-cox method to improve the non-linearity issue. If there's a fanning pattern, apply a function to linearize the pattern as close to linear slope as possible.

Check for two additional conditions where we will plot six predictors in all possible pairs, predictors against respective residuals, and predictors fitted values against residuals. We want to look for scattered points on all plots but the predictors' fitted values against residuals. In fitted values against respective residuals, we'd like to see a linear relationship. If the predictors pairwise plots appear to be linearly related, check for collinearity between predictors by computing the variance inflation factor (VIF). There is collinearity between predictors if it's greater than or equal to 5, it is something worth noting as it would affect our results. Remove predictors one at a time to see which model produces the best VIF value.

Next, identify if there are any high leverage values by the cutoff formula for leverages. Highlighting leverage observations is important because they can also cause potential problems with the model. Check for influential observations by computing three different measures of values: Cook's Distance, Difference in Fitted Values, and Difference in Betas. Using formulas provided in lecture slides, determine if observations of predictors in our model are leverage, outlier, or influential points. We won't remove any observations based on results of this step but only remember their influences on the model and discuss later in conclusion.

Conduct F test to determine if at least one predictor is linearly related to the response. Compute F statistics incorporating residual sum of squares (RSS) and regression sum of squares (SSreg). Compare the value of the F statistic with the p-value to make conclusions.

If we fail to reject the null hypothesis, conduct a partial F test on every predictor to understand further which predictor is better at explaining variations in the response variable. Compute RSS of dropped model and compare with RSS of original model size to determine if predictor should be removed from the model.

Conduct T test on remaining predictors to confirm if predictors affect response variations. Use T-statistic to make the final conclusion to the model.

Results

The exploratory data analysis results showed there are some skewed relationships between some predictors and the response variable.

On predictor pairwise plots, there were many clusters and linear relationships present. We needed to calculate variance inflation factor as well as performing Box-Cox after selecting a more appropriate set of predictors.

After calculating the variance inflation factor, we eliminated all factors but *New HIV infections*, *Birth Assisted by Skilled Professionals,* and *Laws and regulations that guarantee access to sexual and reproductive health care, information and education (percent)*. We checked model assumptions and additional conditions once again. There was a minor fanning pattern between predictors so the response variable was square-rooted in order to linearize the pattern as much as we can. There were also clusters between predictors plots so we performed Box-Cox to fix the non-linearity. From the following graph, we can see that mod 13 (the new model with three predictors) is more linear than mod 1 (the original data with all predictors).

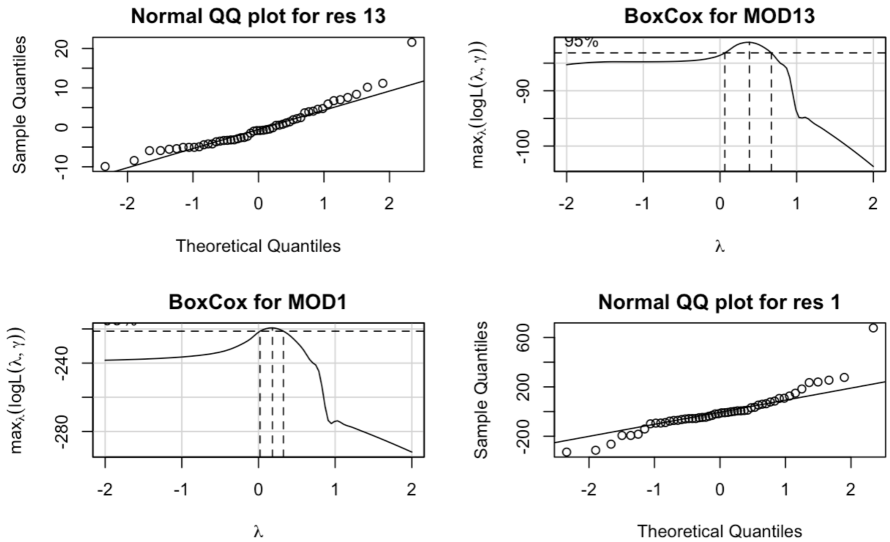

Figure 1 - Comparison of New and Original Datasets

After calculating Cook's Distance, DFFITS, and DFBETAS, we've come up with a few observations that are outliers and influential points.

Cook's Distance	DFFITS	DFBETAS
N/A	1 17 18 21 29	1 17 18 21 24 25 29
Outlier		
43		

Figure 2 - Outliers, Leverages, and Influential Observations

The result of the F test statistic is 22.59 using the new model with three predictors. This tells us that at least one of the three predictors is linearly related to the square rooted *Maternal Mortality Rate*. Next we conducted a partial F test to test if every predictor influences the response. We repeated the test three times because we had three predictors. The null hypothesis is the predictor not influencing the response and the alternative hypothesis is the predictor is influencing the response. The p-value of the test is 0.05. We see that every partial F statistic in the table below is all smaller than the p-value. Therefore, we can reject the null hypothesis and say that all predictors are responsible for the variations in the response.

	Birth attended by professionals	New HIV Infections	Sexual and Reproductive Information Rate
Partial F statistic	0.0103709	0.02260137	0.0227346

Figure 3 - Partial F statistic of predictors

Lastly, we conducted the T-test with a 95% confidence interval. The null hypothesis is the predictor not related to the response variable and the alternative hypothesis is the predictor linearly related to the response variable. After running the code, we arrived at the following t-value of each predictor variable:

	Birth attended by professionals	New HIV Infections	Sexual and Reproductive Information Rate
T-statistic	1.476835e-09	2.103660e-01	1.981997e-01

Figure 4 - T statistic of predictors of T test

We've set the p-value as 0.05. If the T-statistic is greater than the p-value, we will not reject the null hypothesis, however, if the T-statistic is not greater than the p-value, we will reject the null hypothesis which means the predictor is linearly related to the response. Among the three T-statistics, only *Birth attended by professionals* has one smaller than p-value which means it is linearly related to the response.

Discussion

The conclusion of the study is that birth attended by professionals' proportion is influential on the maternal mortality rate. This conclusion does align with the data mentioned in the introduction section. Even though the conclusion matched, other predictors that I thought were relevant to the maternal mortality rate based on my research were eliminated because of collinearity and other issues.

There were many limitations in the study. To begin with, the dataset is not a representative of the same population as the studies I reviewed in the introduction. The result of the model is only representative of the population. The values could be different and lead to a different result. Another limitation is there are many empty values in the dataset and we removed a lot of the columns because they have one missing value. The missing values may have greatly impacted the result. One other limitation is the data collection method and population sample are too different to produce very similar results.

“Maternal Mortality Rates and Statistics.” UNICEF DATA, September 20, 2021. https://data.unicef.org/topic/maternal-health/maternal-mortality/.

Stokoe, U. “Determinants of Maternal Mortality in the Developing World.” The Australian & New Zealand journal of obstetrics & gynaecology. U.S. National Library of Medicine. Accessed December 17, 2021. https://pubmed.ncbi.nlm.nih.gov/1872781/.

Tikkanen, Roosa, Munira Z Gunja, Molly Fitzgerald, and Laurie Zephyrin. “Maternal Mortality and Maternity Care in the United States Compared to 10 Other Developed Countries.” CommonWealth Fund, November 18, 2020. https://www.commonwealthfund.org/publications/issue-briefs/2020/nov/maternal-mortality-maternity-care-us-compared-10-countries.

Zureick-Brown, Sarah, Holly Newby, Doris Chou, Nobuko Mizoguchi, Lale Say, Emi Suzuki, and John Wilmoth. “Understanding Global Trends in Maternal Mortality.” International perspectives on sexual and reproductive health. U.S. National Library of Medicine, March 2013. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886625/.