Socioeconomic determinants of Child Mortality Rate in Nepal

https://github.com/anind99/SocioEconomicChildMortality

Anindya Auveek

03 April 2022

Abstract

The society we live in is not fair. The livelihood of humans depends on their socioeconomic circumstances. While advancement in science has reduced this gap in livelihood regarding healthcare; there remains a stark contrast between those in higher and lower income classes. A key indicator of livelihood regarding health is infant mortality rates. Studying this, the paper by Schell et al. (2007), found that socioeconomic factors can account for 92% of the variability in Infant Mortality Rates. Consequently, in this paper I explored the effect of socioeconomic factors on child mortality rates specifically within the country of Nepal. I hypothesized that the factors associated with higher income will lead to lesser mortality rates. The results confirmed by belief. However, the data I used was not sufficient to conduct an in-depth analysis. Future studies should use extensive data on the individual level to determine as many socioeconomic factors that are related to child mortality rates as possible.

Introduction

Throughout the last few centuries humanity has made leaps in scientific advancements. This has allowed society to develop into the productive and intricate system that we live in. More importantly it has improved the quality of life for humans. Previously difficult to attain resources such as food, water, heat, and shelter are now readily available. However, one of the greatest impacts of scientific development for humans was medicinal advancement and its accessibility. We, as a society, have been able to move past previously devastating illness such as the plague and flu; ultimately increasing the human life span.

While the advancement of medicine has improved life in general for humans, unfortunately it has not had the same effect for everyone. Although a privileged individual may want to deny it, the society that we live in is unfair. For the purpose of increasing capital, countries and people have an unequal income distribution. On the micro level, for lower income countries this has an effect their living standards and other socioeconomic factors. These socioeconomic factors in turn also effect the health of the population. Consequently, in this paper I will study the effect of socioeconomic factors in the mortality rate of infants in Nepal. Nepal being a lower income country will allow us to predict distal factors that may affect individuals who lack financial security.

In the paper by (Schell et al. 2007), they have mentioned that for the purpose of reaching development goals, influential bodies advocate for resource allocation towards the health sector. They state that a major identifier of population health is Infant Mortality Rate (IMR). The WHO suggested that increased spending in the health sector will show a significant reduction in IMR. However other studies have found that socioeconomic progress (nutrition, housing, hygiene, education, gender

equality, and human rights) are more important. Hence, in their paper (Schell et al. 2007) set out to determine the impact of 5 specific socioeconomic variables (GNI/capita, poverty rate, income equality, young female illiteracy rate) on the variation of national IMR at the global level. While gross national income (GNI) predicts IMR, it doesn't provide policy makers clear guidance on how to reduce mortality.

The result of the finding show that 3 socioeconomic variables predicted 92% of the variation of national IMR the global level. Overall, as expected, GNI was the strongest determinant. However, this doesn't provide direct reasons why infant mortality is caused. (Schell et al. 2007) also found the young female illiteracy rate played a major part in the IMR of lower- and middle-income countries; most effective in the former. In addition, income distribution played a major part in middle income countries. High income countries showed no significant association between socio-economic factors and IMR; emphasizing the unjust fact that income is a major determinant of health.

(Schell et al. 2007) concluded that contradictory to WHO's statement, crude macro measurements such as public health spending do not have an independent effect on health gains. Specific factors such as female illiteracy contribute more to child survival. To extent this study, I used data collected by the ministry of health in Nepal (Pradhan and Govindasamy., n.d.) to determine the effect of socioeconomic factors on health specifically in Nepal. The data contains the MORTALITY RATE of infants of different age groups and is aggregated by socioeconomic factors. I hypothesized that there will be a significant effect by socioeconomic variables on child mortality rates in Nepal. More specifically, socioeconomic factors associated with lower income will lead to higher child mortality.

The results suggest that in Nepal there is indeed an inequality in MORTALITY RATE between individuals with different socioeconomic circumstances. While the data I used doesn't capture many factors on an individual level, it provides an initial look on gap in health between people within Nepal. Being a lower income country, the findings from Nepal can also potentially be extrapolated to other countries who are in a similar economic position.

Data

The data I used is collected by the Ministry of Health in Nepal (Pradhan and Govindasamy. (n.d.)). It was gathered through a two-step stratified sampling process. First regions in Nepal were chosen based on if they were considered rural or urban. Secondly, within those regions, subdivisions were given probabilities base on their population size, and 25 households were chosen. The table I analysed specifically is an aggregation of mortality of infants in different age groups and socioeconomic circumstances. Further description of the data is available in the provided in the datasheet located in the appendix.

The software I used to conduct analysis was R (R Core Team 2020) along with following packages: kableExtra (Zhu 2021), reshape2 (Wickham 2007), ggplot2 (Wickham 2016), tidyverse (Wickham et al. 2019), and haven (Wickham and Miller 2021).

Important Variables

- InfantAgeGroup: This captures the age group that the infant has passed away was in. The data is inherently grouped by the ages of infants. So this is an in important variable for context.
- Urban: The mortality rates of infants in urban regions.

- Rural: The mortality rates of infants in rural regions.

The above 2 variables are important because they indirectly capture the socioeconomic differences between types of residence. For e.g., people urban regions may have better access to health care, and generally be wealthier than people in rural regions.

- Primary: The Mortality rates of infants whose mothers have graduated primary school.
- Secondary+: The Mortality rates of infants whose mothers have graduated secondary school or higher.
- No Education: The Mortality rates of infants whose mothers have not graduated primary school.

The above 3 variables indicate the literacy rate of the mothers of the infants. This is an important socioeconomic factor related to MORTALITY RATES as discussed previously.

- Mountain: The Mortality rates of infants in the mountainous regions of Nepal.
- Hill: The Mortality rates of infants in the hilly regions of Nepal.
- Terai: The Mortality rates of infants in the terai (flatland) regions of Nepal.

The data contains mortality rates grouped by different ecological regions specifically to Nepal. There are different socioeconomic circumstances (such as accessibility to health care, education or water) that may play a part in these regions and therefore these variables are important as well.

- Eastern: The Mortality rates of infants in the eastern regions of Nepal.
- Central: The Mortality rates of infants in the central regions of Nepal.
- Western: The Mortality rates of infants in the western regions of Nepal.
- Mid-western: The Mortality rates of infants in the mid-western regions of Nepal.
- Far.western: The Mortality rates of infants in the far-western regions of Nepal.

The data contains mortality rates grouped by different geographical regions specifically to Nepal. There are different socioeconomic circumstances (such as the income and education) that may different in between regions and therefore these variables are important as well.

- No antenatal or delivery care: The mortality rates of infants who have received no antenatal or delivery care.
- Either antenatal or delivery care: The mortality rates of infants who have received either antenatal or delivery care.
- Both antenatal and delivery care: The mortality rates of infants who have received both antenatal and delivery care.

These variables are directly associated with infant health, but the data is incomplete for these. Therefore, we will not be analysing these for this study.

To follow, I will be providing the mean mortality rates of the different variables along with the comparison of variables with each other.

Table 1: Mean Mortality Rates of Selected Socio-economic Groupings

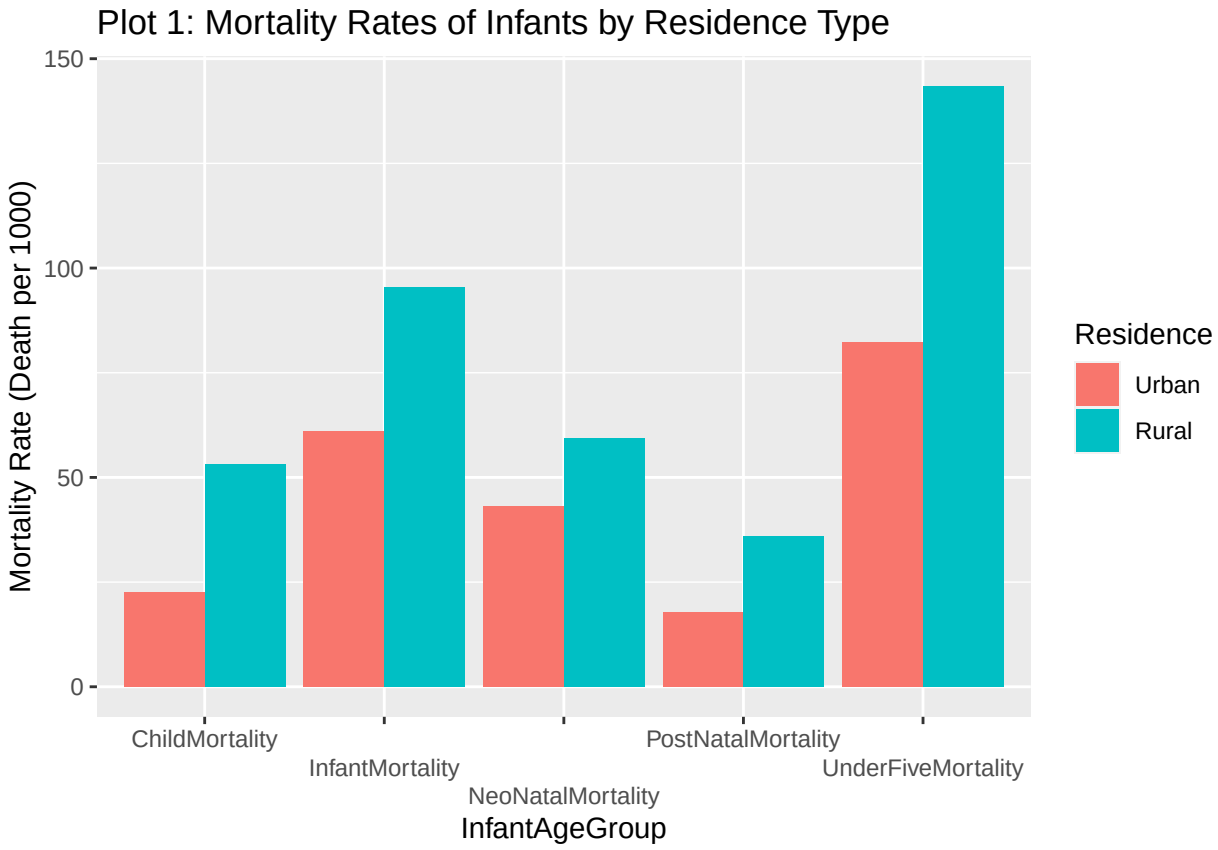
SocioEconomicGroup	MeanMortality
Urban	45.4
Rural	77.4
Mountain	113
Hill	69
Terai	74.8
Eastern	61.6
Central	73.2
Western	65
Mid-western	95.7
Far.western	98
Primary	56.1
Secondary+	35
No Education	80.1
No antenatal or delivery care	53.3
Either antenatal or delivery care	33.3
Both antenatal and delivery care	NA

Table 1 contains the mean mortality child rates for each variable. We can see at glance that there are indeed differences in mortality rates between types of residence, ecological regions, and geographical regions.

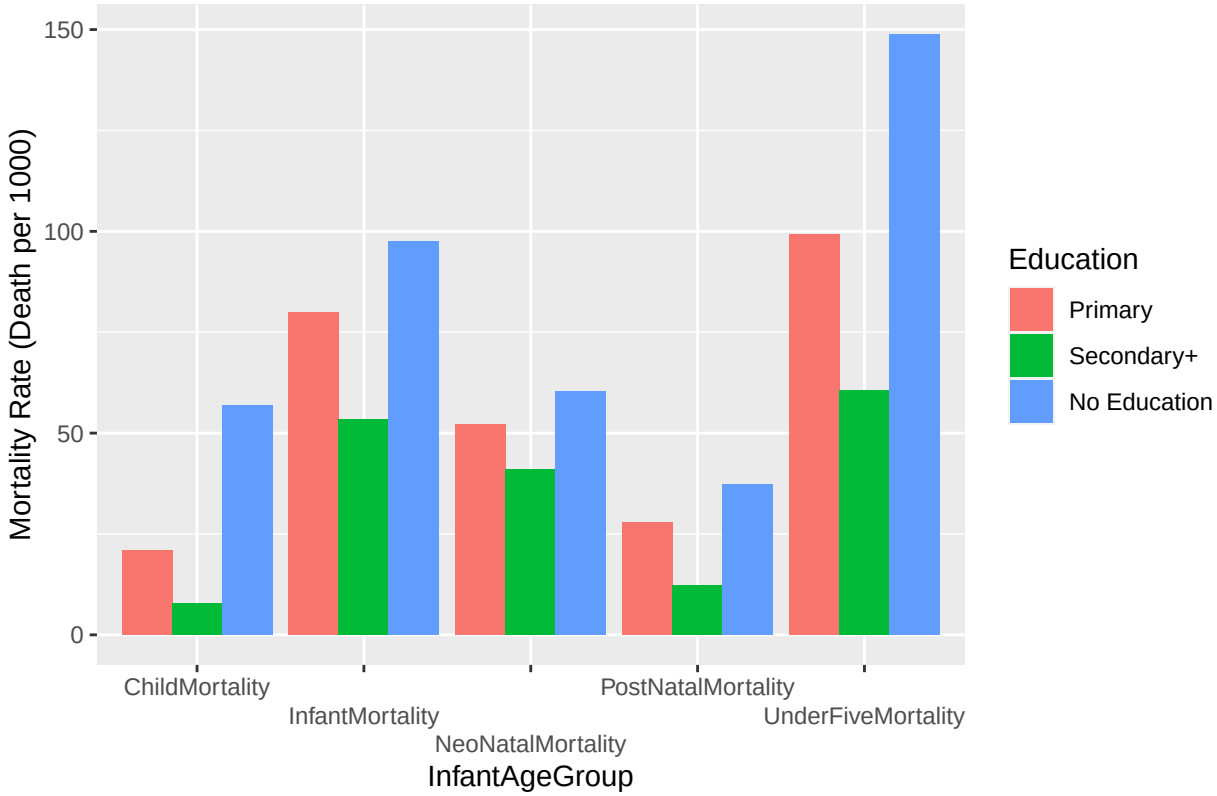
- Urban regions have lower mortality than rural ones.
- Children whose mothers have a higher level of education have lower mortality rates.
- Hilly regions have lower mortality than Terai regions, who have lower mortality than mountainous regions.
- The mortality of geographical regions is ascending in this order: Eastern, Western, Central, Mid-Western, Far Western.

Whether these differences are only in the mean mortality rate or for all age groups must be analysed further.

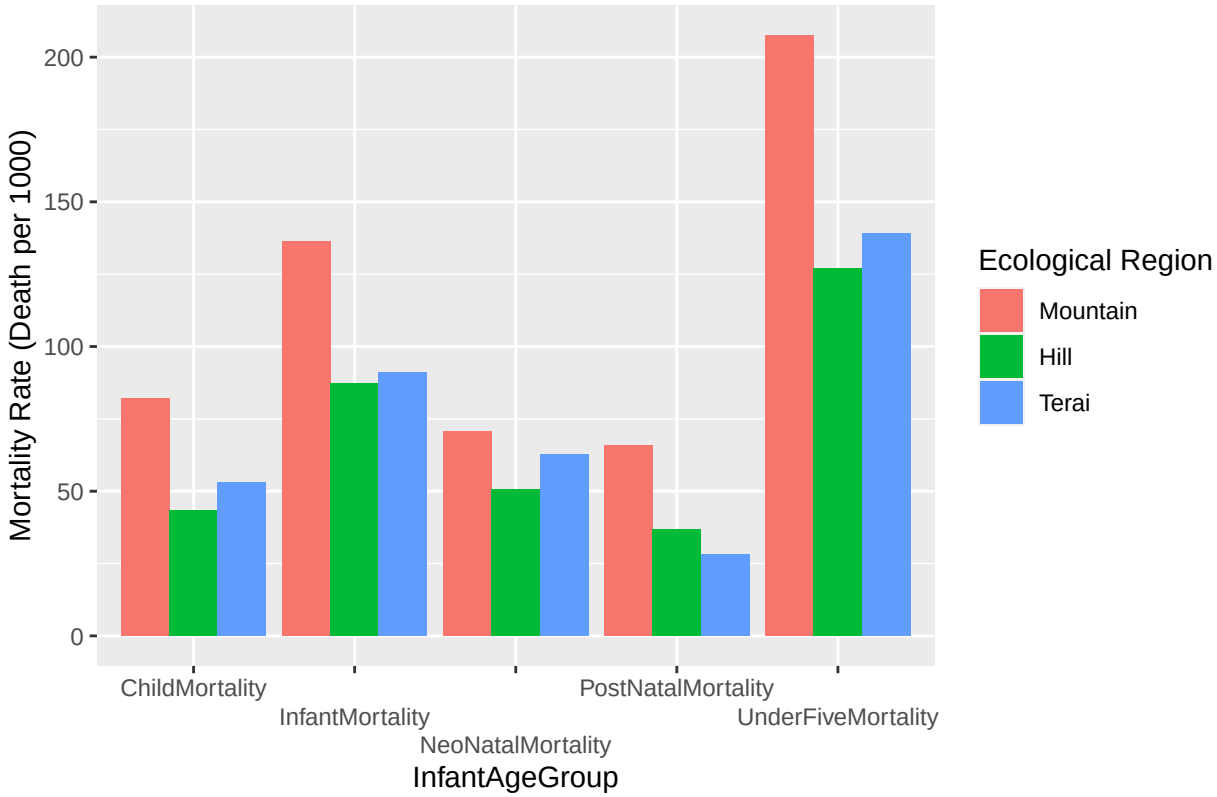
Supplementing the mean table, I will be displaying the distribution of mortality rates by age group of the variables in comparison with each other.

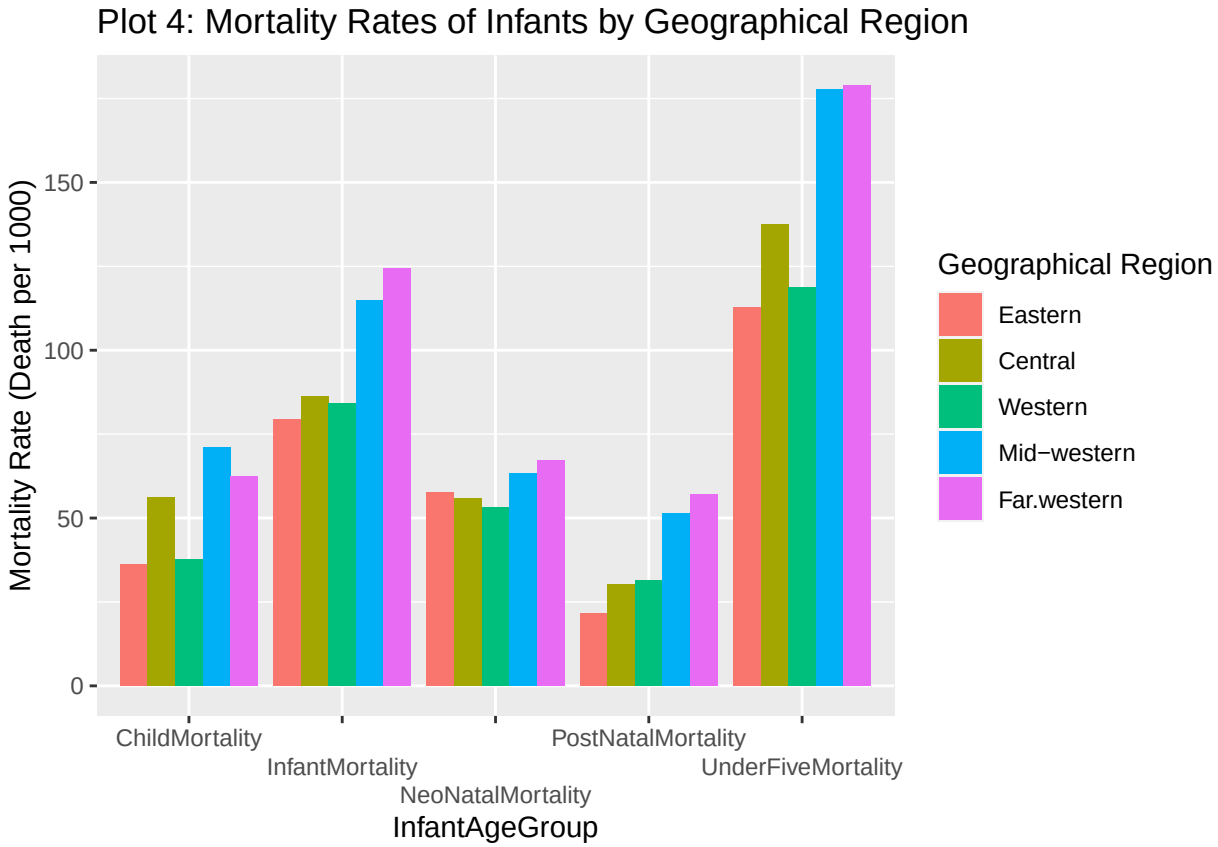


Plot 2: Mortality Rates of Infants by Education Level



Plot 3: Mortality Rates of Infants by Ecologicial Region





Plot 1 shows the mortality rate of different age groups by residence type. We can see that in most age groups the MORTALITY RATE in rural regions is higher. However, the mortality rate of the "under 5" age group is significantly higher in urban regions. This is an outlier.

Plot 2 shows the mortality rate of different age groups by the mother's education level. We can see that for the lower age groups, MORTALITY RATE is lowest if their mothers have a secondary or higher level of education. However interestingly, this is not found for children in higher age groups. Children who have mothers with "no education" surprisingly have the lowest mortality rates if they are in the under 5 age group.

Plot 3 shows the mortality rate of different age groups by Ecological Region. We see that there are no generalizable differences in the MORTALITY RATEs of age groups grouped by their ecological region. Overall mountainous regions seem to have the highest, however for the "infant" age group, it has the lowest mortality rates.

Plot 4 shows the mortality rate of different age groups by Geographical Region. We can see that in general eastern, western, and central regions have lower mortality rates than mid-western and far-western regions. However surprisingly for the infant age group, they have a higher rate. This is similar finding to the grouping by ecological region and may be due to the same factors.

After looking at these distributions of mortality rates by age groups and comparing the socioeconomic variables, we must determine whether these differences are significant. This is what I will do in the results section.

Results

To identify the statistically significant effect of socioeconomic factors on mortality rates I will be conducting t-tests between the variables. Since the data is limited, I will be using one-side tails with a significance level of 90% to determine if the mortality rate of one variable is lower or higher than another. A p-value of less than 0.1 will contradict the null hypothesis and confirm the alternate hypothesis that 1 variable has a higher mortality rate than another.

Table 2 contains the results of these tests. The interpretation of the data in the table is as following:

- Null Hypothesis: There are no differences between the mortality rates X1 and X2 by age group.
- Alternate Hypothesis: The mortality rates of X1 is different than X2 by Type (Less or Greater).
- If P-Value is less than 0.1, then we accept the alternate hypothesis.

Mothers Education Level

1. We see that the mortality rates of children with mothers that have an education level of secondary or higher is significantly lower than the mortality rates or infants with mothers that have no education.
2. However, there is no significant differences in the mortality rates of other education levels. It is important to note that this may be due to the lack of data as the p-values are still close to 0.1.

Type of Residence

We see that mortality rates in Urban residences are significantly lower than mortality rates in rural residences. These indicate the socioeconomic differences between rural and urban regions result in a health advantage for people living in the latter.

Ecological Region

1. There are no significant differences in mortality rates between ecological regions.
2. However, it is important to note that this may be due to lack of individual data, as the p-values of mountainous vs hill and terai are close to 0.1. The mortality rates of mountainous regions may indeed be higher due to socioeconomic factors.

Geographical Region

1. There are no significant differences in mortality rates between geographical regions.
2. However, it is important to note that this may be due to lack of individual data, as the p-values of Eastern and Western vs Mid-western and Far-western are close to 0.1. The mortality rates of Mid-Western and Far-Western regions may indeed be higher due to socioeconomic factors.

Discussion

The society that we live in is not fair. Humans who are born into a live of socioeconomic privilege are much more equipped to succeed in life than humans who are not. The scientific advancements made in medicine closed some these gaps in livelihood between humans of different socioeconomic

Table 2: T-Test of Mortality Rates by Socioeconomic Factors

X1	X2	Type	P-Value
Urban	Rural	Less	0.0999
Mountain	Hill	Greater	0.107
Mountain	Terai	Greater	0.144
Hill	Terai	Less	0.413
Primary	No Education	Less	0.181
Secondary +	No Education	Less	0.045
Secondary +	Primary	Less	0.145
Eastern	Western	Less	0.442
Eastern	Central	Less	0.323
Eastern	Mid Western	Less	0.132
Eastern	Far Western	Less	0.122
Western	Central	Less	0.373
Western	Mid Western	Less	0.156
Western	Far Western	Less	0.144
Central	Mid Western	Less	0.235
Central	Far Western	Less	0.217
Mid Western	Far Western	Less	0.474

groups. One specific benefit from these advancement over the last century was the steady reduction of infant mortality rates. However there remains a gap between higher and lower income countries.

In the paper by (Schell et al. 2007) they have found that 3 socioeconomic variables predicted 92% of the variation of national IMR at the global level. These factors re GNI, Income Distribution and Female Illiteracy Rates. In lower- and middle-income countries specifically, they found that Female Illiteracy Rates and Income Distribution played the biggest part in IMR. Suggesting that these socioeconomic factors are either direct- or indirect- causally responsibly for IMRs. They concluded that policy makers should focus on causes like these on the micro level to improve the general health of the nation.

To extend the previous study, in this paper I explored the effect of socioeconomic factors on child mortality rates within the country of Nepal. Nepal being a lower income country itself can provide insight into the aspects of child mortality in other lower income countries as well. My hypothesis in the study was that socio-economic factors will indeed influence child mortality in Nepal. Specifically, following the results of the study by (Schell et al. 2007), I predicted that the education level of the mothers of children will have a significant impact on child mortality.

In order to study this, I used data collected by the Nepal ministry of health in 1996 (Pradhan and Govindasamy., n.d.) and created a dataset containing the mortality rates grouped by different socio-economic circumstances (variables).

I began the analysis by looking at the mean child mortality rates of different socioeconomic variables. These means are essentially capturing the weighted average of child mortality rates; with each age group being give the same weight. Therefore, they do not represent the actual population mean of mortality rates for each socioeconomic variable.

Several interesting findings can be seen from these means. Aligning with my hypothesis, the higher the education level of the mother the lower the mortality rate (i.e., Secondary+ < Primary < No Education). In addition, urban regions also have a lower mean child mortality rate than rural ones. This also aligns with my hypothesis since urban regions generally have a higher income than rural ones. Lastly, different geographical and ecological regions also had different mortality. Although it cannot be conclusively said whether this aligns with my hypothesis, the income distribution in these regions may be different from each other. Thus, resulting in different socioeconomic factors which affect child mortality.

Probing deeper into the relationship between the variables I plotted the mortality rate of different age groups by groups of variables.

The first group being residence type. The plot shows that urban regions do indeed have a lower mortality rate than rural regions almost all age groups, aligning with my hypothesis. However, the mortality rate in the "under 5" age group is significantly higher in urban regions. This may be due to non-socioeconomic factors. The general reasons which lead to child mortality for lower ages may not apply to higher age groups.

The second grouping of variables I plotted were the mother's education level. As hypothesized, children with mothers who have a higher education level have lower mortality rates. However, like the finding in the residence type plot, the higher age group of "under 5" do not share the finding.

The other two grouping of variables I plotted were ecological region and geographical region. In these groupings I could not see any significant differences of child mortality rates. For example,

overall mountainous regions seem to have the highest mortality rates, however for the infant age group they have the lowest.

While simply looking at the differences in mortality rates between different socioeconomic variables provided me an initial insight. It did allow me to make any conclusions. Therefore, to determine whether the actual differences between the groupings of variables were significant, I used t-tests.

The findings for the t-tests of the mother's education level show that mothers who have a secondary or higher level of education result in a significantly lower child mortality than mothers with no education. However, no significant findings were shown between primary level of education and no education. The latter lack of finding may be due to the limitation of the data; the table only contains grouped information and consists of only 5 rows. The former confirmation of the hypothesis can be explained through various attributes related to having a higher education. Indirectly, if you have a higher level of education, you are more likely to be in a higher economic group. Thus, having better access to medicine and other direct preventers of child mortality.

To add, the findings for the t-tests of the type of residence show that urban residences do have significantly lower child mortality rates than rural residences, aligning with my hypothesis. As previously mentioned, this may be due to the socioeconomic differences between urban and rural regions. However, a plethora of other socioeconomic factors not available in this dataset may also be responsible.

Lastly, the findings for the t-tests of the ecological and geographical regions do not show any significant results, contradictory to my hypothesis. Nevertheless, this can also be attributed as a limitation of the dataset. Since the data does not contain enough records, and is not on the individual level, t-tests will not be able to capture the true distribution of the variables. Therefore, we cannot conclude whether the lack of findings is true.

Overall, the statistical analysis of the variables has been in line with my hypothesis. However, a major limitation of this study is the lack of extensive data on the individual level. Since the table I extracted data from was already grouped by age-group, the actual distribution of the sample could not be used for any t-tests.

Another limitation resulting from the dataset is that there was only a handful of socioeconomic variables recorded. A future study should attempt to capture a larger plethora of factors such as income, education, employment, community safety, and social supports and analyse their relationship with child mortality rates. Sampling should also be stratified to accommodate enough subjects within each socioeconomic circumstance. It is important to have a larger array of variables in order to determine which variables are independently responsible for the variation in child mortality (one variable may become obsolete when another is taken into account).

Bibliography

Pradhan, Ram Hari Aryal, Ajit, and PavalavaIli Govindasamy. n.d. *Nepal Family Health Survey 1996.* http://dhsprogram.com/pubs/pdf/FR78/FR78.pdf.: Nepal Family Health Survey 1996.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schell, Carl Otto, Marie Reilly, Hans Rosling, Stefan Peterson, and Anna Mia Ekström. 2007. "Socioeconomic Determinants of Infant Mortality: A Worldwide Study of 152 Low-, Middle-, and High-Income Countries." *Scandinavian Journal of Public Health* 35 (3): 288–97. https://doi.org/10.1080/14034940600979171.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

Appendix

Data Sheet

Motivation.

1. For what purpose was the dataset created? The dataset was created to capture information regarding the effect of socioeconomic factors on infant mortality in Nepal.
2. Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)? It was created by the Ministry of Health (Nepal).
3. Who funded the creation of the dataset? It was funded by the Nepali Government.

Composition.

1. How many instances are there in total (of each type, if appropriate)? There are 5 instances of Mortality rates for all 16 socio-economic categories (i.e. 80 instances total).
2. Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? The dataset is a summary of a sample of the population (all instances). The sample size is 8429.
3. What data does each instance consist of? Mortality Rate (Deaths per 1000 individuals).
4. Is there a label or target associated with each instance? Yes, for example the mortality rate of
5. Is any information missing from individual instances? Yes this specific data set does not contain information about individuals, rather it is a group summary based on a socio-economic demography.
6. Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? Yes, the individual instances are linked based on their type of grouping. For example geographical location, Education, etc.
7. Are there recommended data splits (for example, training, development/validation, testing)? No
8. Are there any errors, sources of noise, or redundancies in the dataset? Yes, some of the instances are Not available or have an asterisk or an in a parenthesis. Figures in parentheses are based on 250-499 births; an asterisk indicates that the figure is based on fewer than 250 exposed persons and has been suppressed.
9. Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? It is self-contained.
10. Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? No, it is summary data (not individual).
11. Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? No

People

1. Does the dataset identify any sub-populations (for example, by age, gender)? Yes the subpopulations are divided by socioeconomic group.
2. Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? No

3. Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? No
4. Any other comments?

Collection process.

1. How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? The data was reported by subjects.
2. What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? It was collected through a direct interview with the subjects.
3. If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)? The data was collected through two-step stratified sampling. The sampling strategy was probabilistic. The first step was the allocation 253 PSU (primary sampling unit). 34 were allocated to urban areas, 219 were allocated to rural areas. Then in these areas, each subdivision was given a probability proportional to their population size (number of households). 25 Interviews were completed from each PSU probabilistically.
4. Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)? Workers from the Nepal Ministry of Health.
5. Over what timeframe was the data collected? This data was collected for the year 1996.
6. Were any ethical review processes conducted (for example, by an institutional review board)? No.

People 2

1. Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)? Obtained it through DHS.
2. Were the individuals in question notified about the data collection? No
3. Did the individuals in question consent to the collection and use of their data? The individuals are kept anonymous, so they do not need to consent to each data collection (just the initial).
4. If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? No.
5. Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? No

Preprocessing/cleaning/labeling.

1. Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? Yes, the initial dataset was transposed to make the rows columns as they represented a specific grouping.

2. Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? Yes: raw_data.csv
3. Is the software that was used to preprocess/clean/label the data available? Script/clean_data.r

Uses.

1. Has the dataset been used for any tasks already? It was used to analyse the difference in mortality rates between different socioeconomic groups.
2. Is there a repository that links to any or all papers or systems that use the dataset? No
3. What (other) tasks could the dataset be used for? None
4. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? No
5. Are there tasks for which the dataset should not be used? No Distribution.

People 3

1. Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? Yes, it can be accessed by anyone.
2. How will the dataset be distributed (for example, tarball on website, API, GitHub)? It is available online at DHS.
3. When will the dataset be distributed? It has already been distributed.
4. Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? No
5. Have any third parties imposed IP-based or other restrictions on the data associated with the instances? No
6. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? No

Maintenance.

1. Who will be supporting/hosting/maintaining the dataset? No one
2. How can the owner/curator/manager of the dataset be contacted (for example, email address)? You must contact the ministry of health (Nepal).
3. Is there an erratum? No
4. Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? No
5. If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? No
6. Will older versions of the dataset continue to be supported/hosted/maintained? N/A
7. If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? No