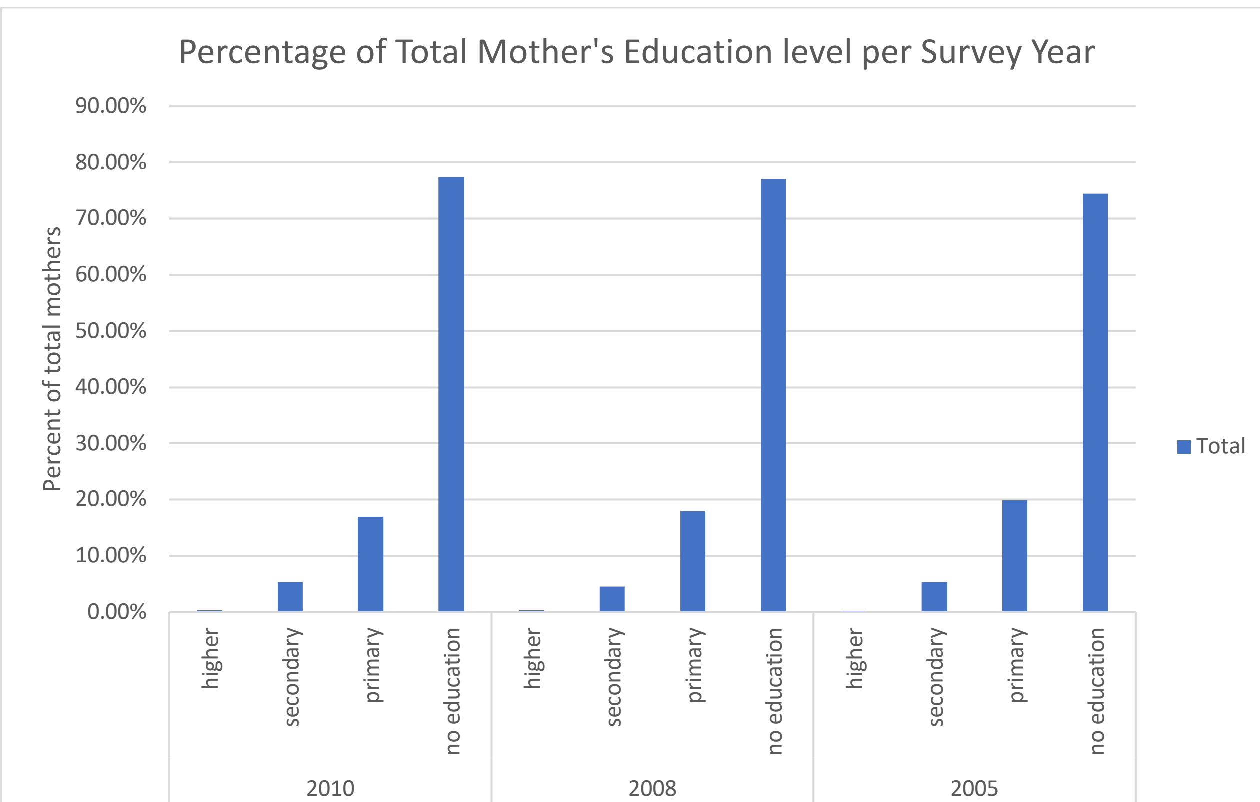
Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

3a)

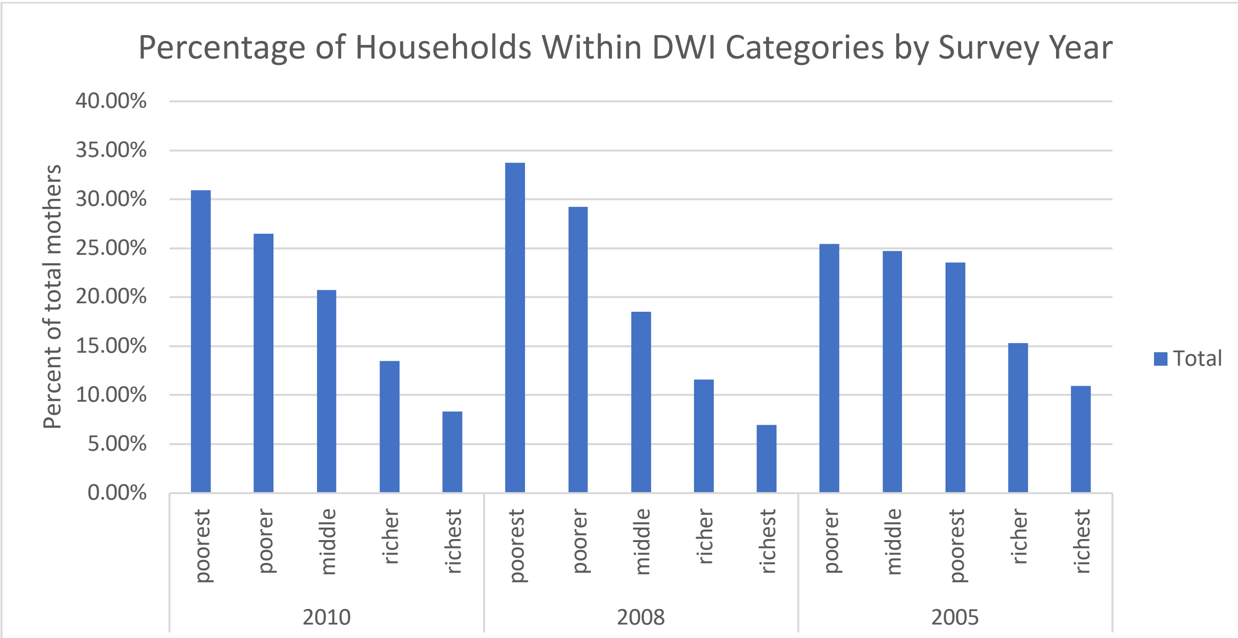
To create the bar chart below that shows the number of mothers within each education level categorized in the “educ_mom”, a pivot table was created to count the totals with respects to each survey year with “educ_mom” and “svy_yr” in rows. The pivot table is then used to create a pivot bar chart where the counts of these values are being displayed in percent of parent row total. The use of percent of parent row total allows the graph below to illustrate the percentage of mothers that have which level per survey year. In conclusion, in each survey year there is a consistent trend where the higher a mother is educated, the lower the percent frequency is with respects to the total mother surveyed. But the trend has a steep drop in percentage for educations above “no education”.



Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

3B)

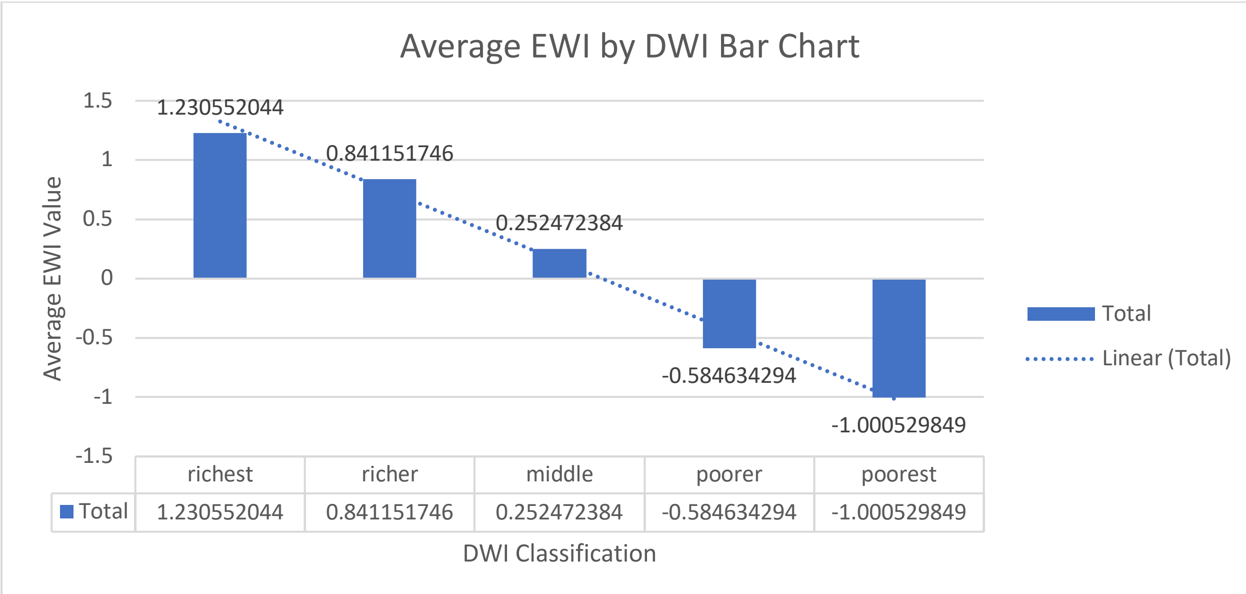


First, I created a pivot table to calculate the count of mothers that were assigned each DWI category by putting "svy_yr" and "DWI" into rows and from there created a pivot chart. The Y-axis of the bar chart is in percentage of parent row total which allows us to view the percentage of households that are within each DWI category by survey year. The figure above illustrate a trend of which there is a higher percentage of households that are considered poor which decreases in proportion as it goes to towards the richest.

3C)

Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4



First the qualitative categories that the DWI uses, had to be converted into a quantitative measurement. This is due to the inability to directly measure qualitative and quantitative measurements. To have a more clear comparison, I gave numerical values to each of the categories (Richest= 1, Richer = 2, middle = 3, Poorer = 4, and Poorest= 5) using an "ifs" function to measure both indices with the same units. Then, I did a pivot table with DWI or DWI quantified as the column; and an average corresponding EWI value to find the possible relationship between the two with a pivot bar chart for the figure above.

What can be seen visually is that there is a direct relationship of, as you go richer in DWI value there is a larger EWI value. The same can be said if you go poorer and reach a negative number for poorest. This visual indication was shown through the trend line that illustrates this correlation between the two methods of measuring income. Both methods are seen to be important to categorizing larger samples for DWI, and EWI for smaller sample sets.

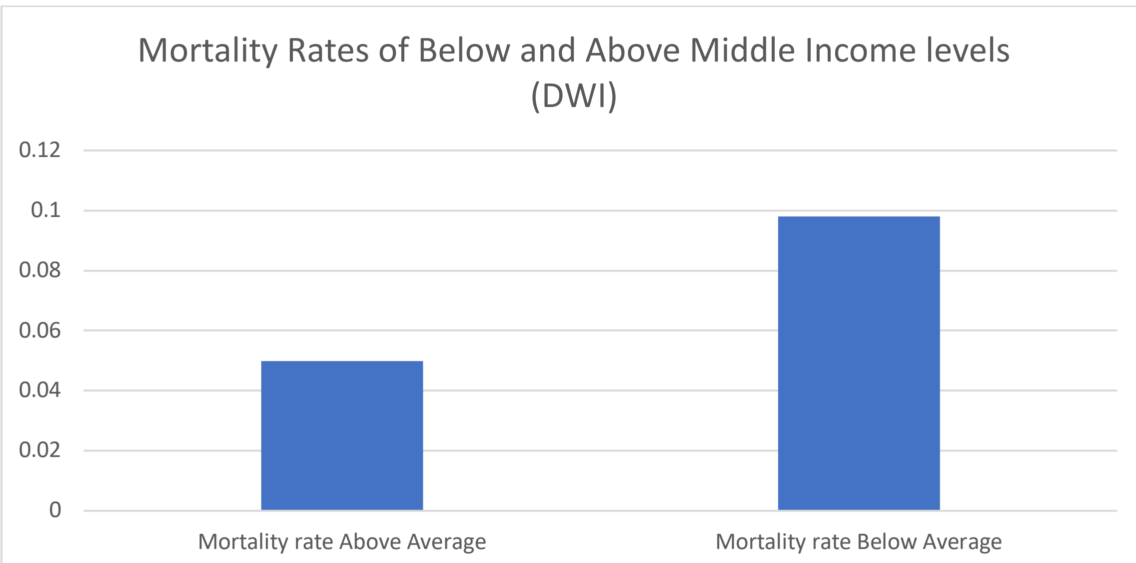
3D)

I created bar chart figure below to illustrate the morality rates of income levels above and below the Middle income values. I utilized the DWI to find the mortality rates because "middle" is already an estimate for an average income level for a household. The two values above and

Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

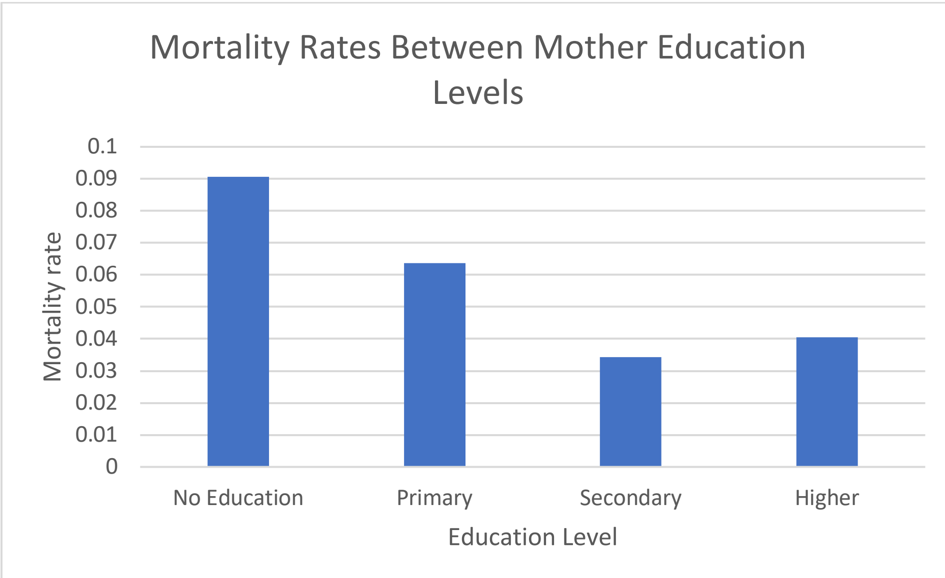
below this "middle" value are used for above and below average incomes. With this in mind, I created a pivot table with counts of births and counts of deaths through the "Age_at_death" and "birth_year" that correspond to these DWI values. Afterwards to find each wealth level's mortality rate by dividing the total amount of people dead and alive by the amount of people who were dead. In conclusion, the mortality rates of above average incomes are less than the mortality rates of below average incomes and shows that mortality rates decrease with an increase in wealth.



Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

3E)



Row Labels	Count of birth_year	Count of age_at_death
higher	173	7
secondary	2802	96
primary	10060	641
no education	43098	3904
Grand Total	**56133**	**4648**

I created a pivot table of the mother's education level, "age of death", and "Birth years". Using the "birth years" and "age of deaths" as proxies for total births and total deaths and their counts in the pivot table, I then divided deaths by births to find the mortality rates of each corresponding education levels. What can be seen is that, other than secondary educated mothers, the mortality rates decrease as the mother has higher levels of education. By finding the totals that correspond with each education level, I can find the mortality rates that correspond as well. The higher-level category may be seen to have a higher rate compared to secondary because its sample size is significantly smaller compared to the rest of the data.

Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

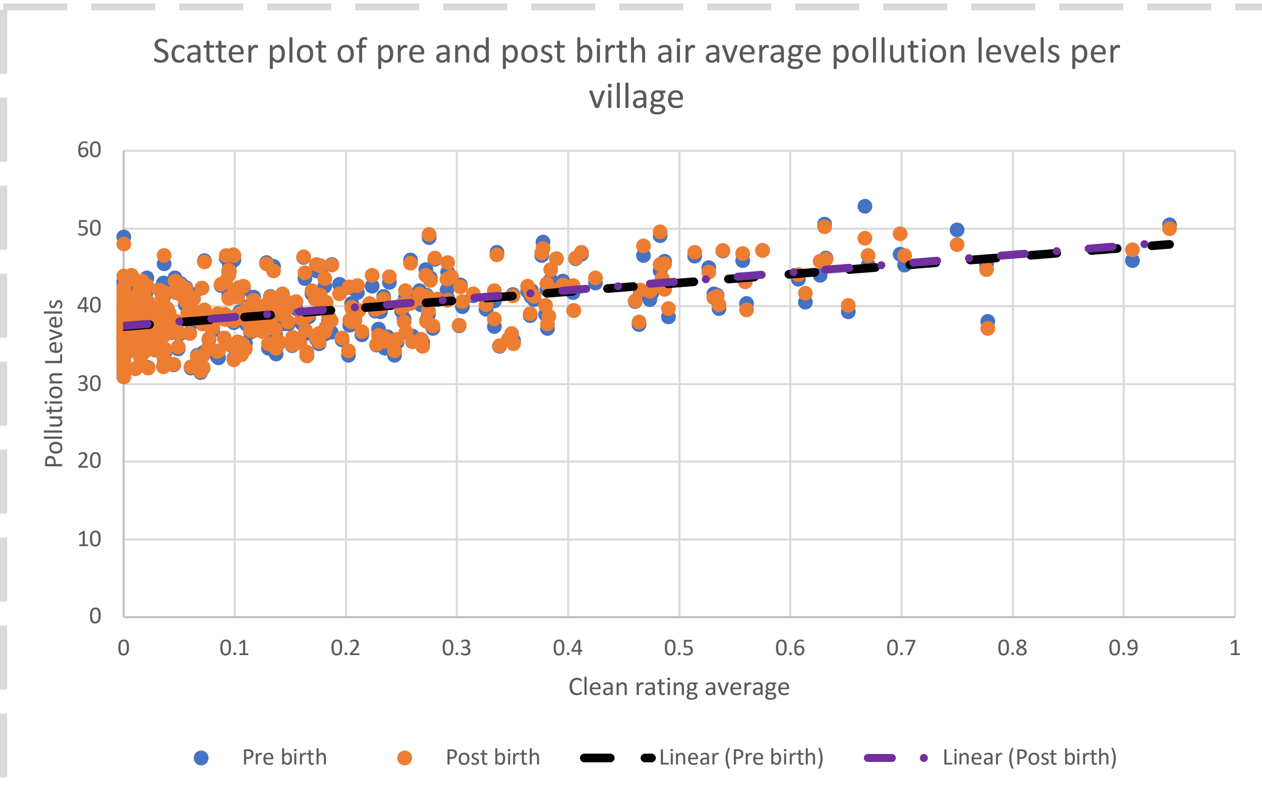
4A)

To see if there is a statistical difference between the mortality rates of homes that have clean cooking or unclean cooking, a null hypothesis was to be if the difference between the true population means of clean households and unclean households is equal to 0. Within the clean cooking data, those with a "1" are considered clean cooking and those with a "0", unclean. The alternative hypothesis, will be that there is a difference between the population means of each. These hypothesis will be written as, $H_o\colon \mu_{clean} = \mu_{unclean}$ and $H_A\colon \mu_{clean} \neq \mu_{unclean}$ with $\mu = mortality\ rates$ of each category. Basically I wanted to see if there was a difference in pollution, pre and post birth, for homes with or without a clean cooking sourse. This test to find if there is a difference or not, requires a two tailed hypothesis test with a predetermined significance alpha level of .05 which will allow us to see with a 95% degree of confidence that if we can reject the null or not. To aggregate the homes with these cooking sources, I utilized village cluster IDs to test all the households in each village. And from this village cluster ID column, of which there are 392 unique villages that were surveyed no matter the survey year, I found each villages' average air pollution before and after birth to create the arrays of data needed for a "T.Test" function that will have a two tailed and unequal variance criteria. The P-value that was calculated was **1.02234E-20**, which we can easily see is lower than our .05 alpha. Therefore we can conclude that we have sufficient information and evidence to say that there is a statistical difference between pre and post birth air pollution levels between homes that have clean or unclean cooking sources. It is important to note that just because we can find a statistical difference between the data, it was calculated to see if there was a difference and not if "clean" cooking was statistically "better" than "unclean" with mortality rates. It may be totally plausible that "clean" households may in fact be worse off than "unclean" homes.

Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

4B



I found the clean rating of each village cluster by averaging out the clean ratings of all the households. The closer the number is to 1, is considered cleaner and cleaner. The opposite can be said as numbers reach 0 for how unclean the average household is in a village. For each of these cleanliness averages, their corresponding in utero and post birth pollution was found through "pm25_pre" for in utero times, and "pm25_post" for post birth situations. Afterwards, I scatter plotted this data for each village, to see if there is an effect of with clean cooking after having birth. Additionally, I placed the trend lines to see if there is movement between cleanliness rating averages and pollution levels emitted. In conclusion, the scatter plot shows that pollution has a higher density grouping in villages that have a lower clean cooking source rating. This is seen in

Leo Do
Professor Teevrat Garg
Quantitative Methods 1
6 December 2019

QM Project Writeup: Questions 3-4

the high-density data placement which is in the very left of the chart. This is important to note because we see on the plot, that homes with “cleaner” cooking sources are emitting on average higher pollution levels. Because of this finding, it shows that correlation does not imply causation that we statistically found in 4a where there is a difference between clean and unclean homes for air pollution. This finding allows us to conclude that, air pollution from the average village that comes from clean cooking may not be the major factor lowering mortality rates but rather other factors.