Frances Sinclair

CIS 241

Dr. Hallenbeck

28 September 2017

A Study on Familial Sizes Indonesia

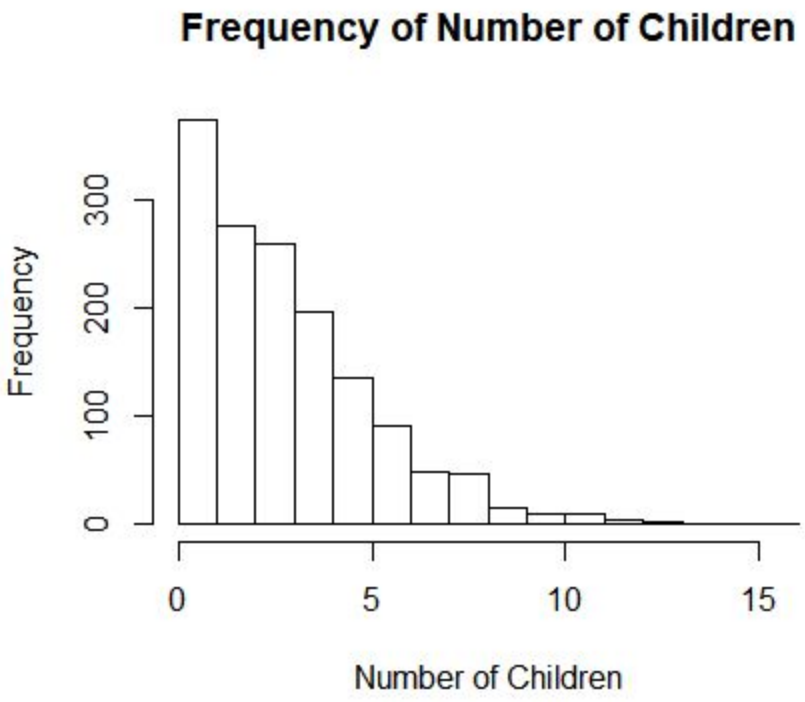
Does education, social status, and religion affect the size of a family? I wanted to see how the life around a family could affect its size. I chose a dataset from a survey in 1987 on Indonesian Contraceptive Prevalence. I chose this dataset because it had a lot of different consistent variables. Some other studies I saw would ask about alcohol consumption and tobacco use and the weight of the child being born, however the questions were not asked consistently across all data points. This would make it harder to study their affect and their prevalence. The formatting of other files also made them significantly harder to use and their variables were not adequately described or explained.

Indonesia is an island with the Indian Ocean to the West and the South Pacific ocean to its East. Since space is limited, overpopulation has been a huge issue. The government had then been promoting having smaller families. The worst of the overpopulation occurs in the larger cities where public transportation is atrocious. The average income of a family in Indonesia is about $14,500 (USD) (1). This is not a lot of money considering the size of some families living there.

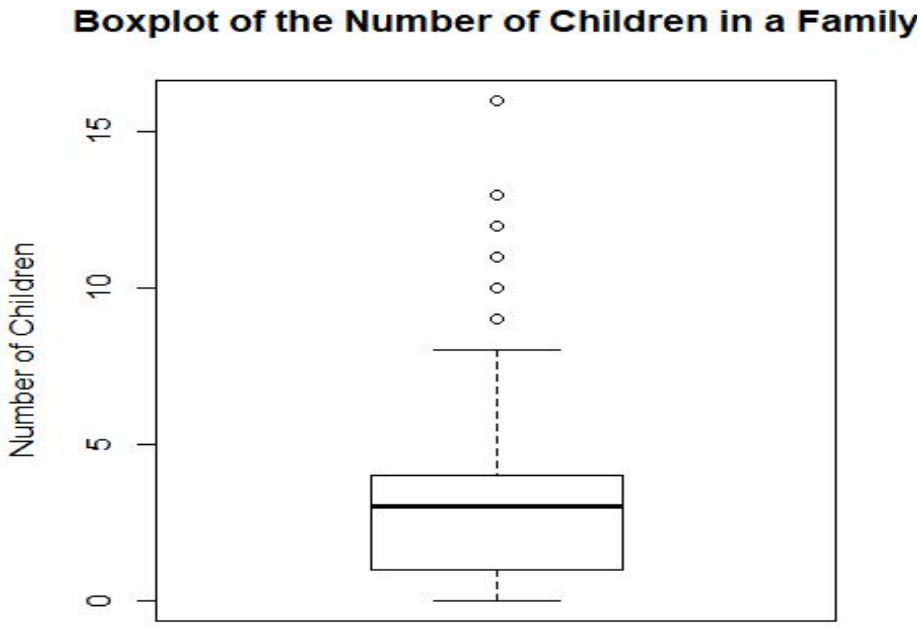
The Dataset observed in this paper is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. The variables studied in this survey were the wife's age, the wife's level of education on a scale of one to four, one being the lowest and four being the highest, the husband's education based on the same scale, their number of children, whether or not the wife identified as islamic, whether or not the wife was currently working, their standard of

living one being the lowest and four being the highest, their media exposure as good or not good, and their contraceptive method use as none, long-term, or short-term.

In the analysis of the data, a distribution of the Number of Children is a good place to start. It demonstrates how many children to a family was most frequent and can also help demonstrate the slope of the graph of the number of children. The most popular number of children was zero. The frequency of the number of families with children above 13 were so low that they are not even visible on the histogram above, except for a straight line. The highest number of children in the survey was 16 children and only one family had that many. Compared to the number of families with no children those with 16 are merely just thin line on the graph. The majority of the families depicted on the graph had somewhere between zero and five children and very few had anything above that. Even less had more than 10 children.

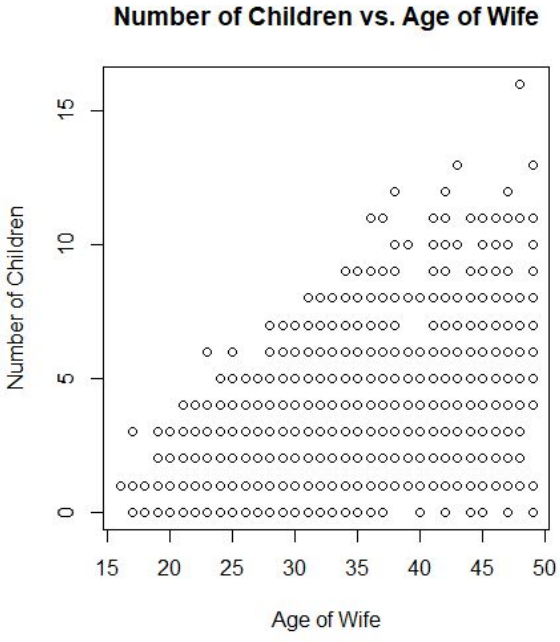


Below is a boxplot of the number of children born to the subjects who were studied. This shows that the median number of children in the families that were surveyed was 3. This was surprising due to the sheer number of families that had zero children. The upper limit for outliers



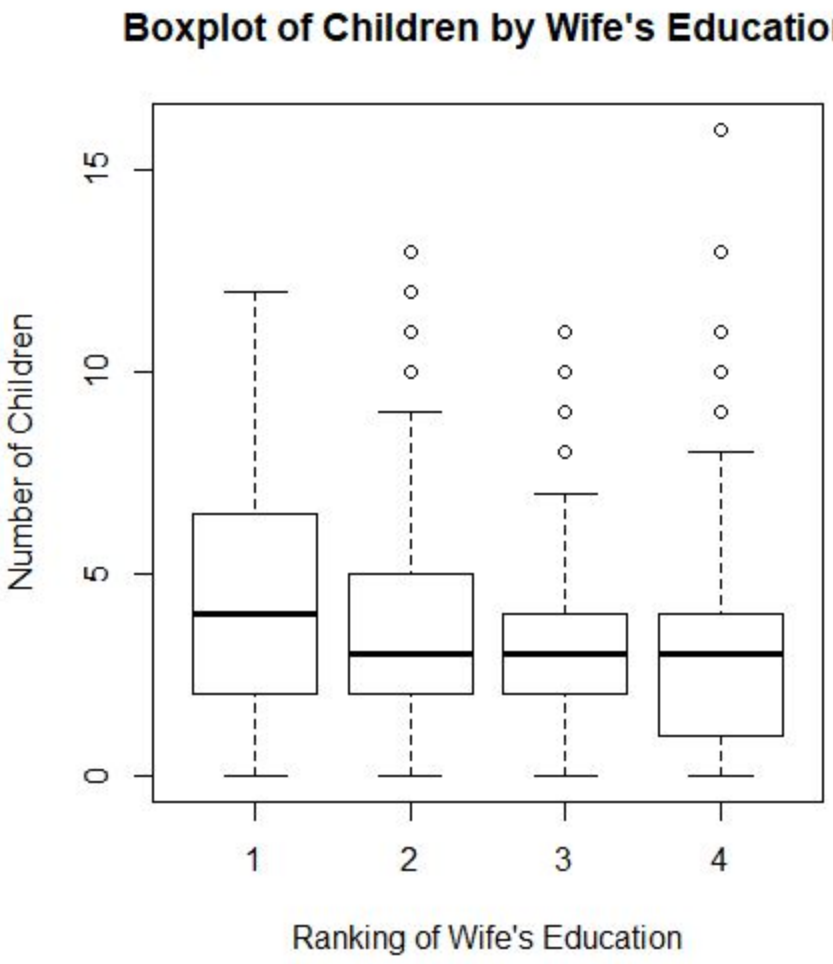
is 8.5 children. The lower limit is 0, since no one can have less than no children (if calculated the same way as the upper limit was calculated, the lower limit was in the negatives). I decided not to take out the outliers, even though they are so far above the turkey line, because they are important in order to see the conditions under which 16 children (the highest data point) and other outlying data points were born.

The following graph is the number of children in a family compared to the age of the wife. The number of children and the age of the wife had the greatest correspondence out of all the variable compared to the size of families. The graphs of education of the parents and media exposure showed no significant correlation with the number of children in a family. The points were evenly spread out through the two options. As the age of the wife increased so did the

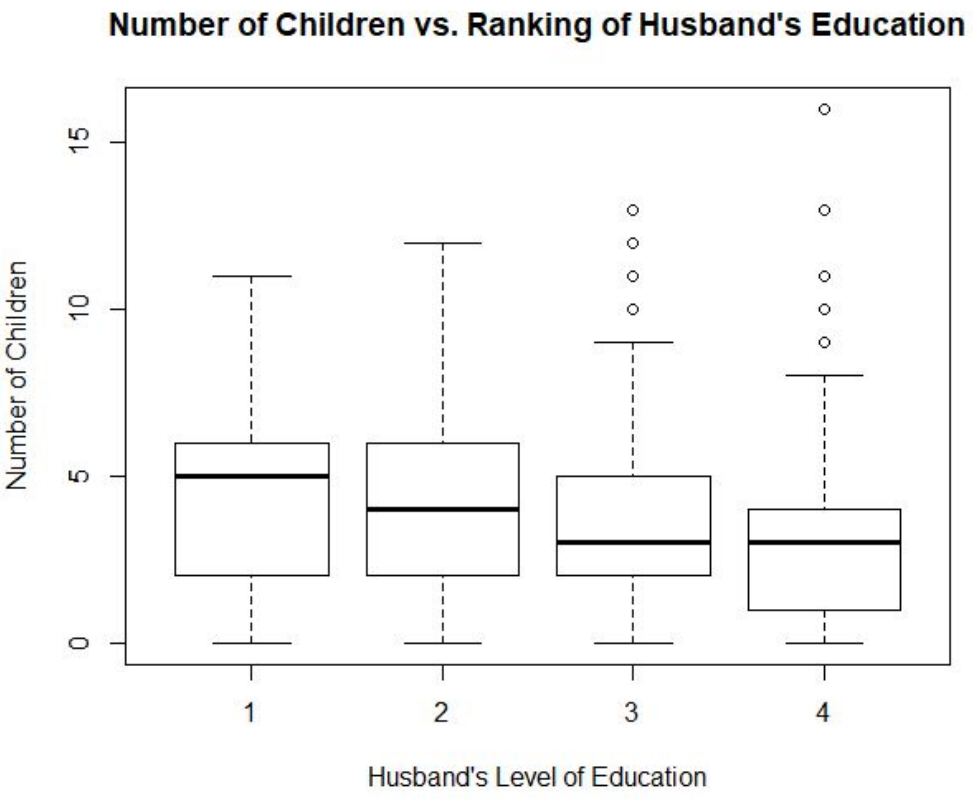


number of children a family had. This makes sense since it takes time to have more children so the women who are 40 or 45 have had decades to give birth and grow the number of children in their families. The graph also shows that they only received surveys from people above the age of 15 and below the age of 50.

In the figure below there is boxplot of the number of children compared to the wife's level of education on a ranking of one through 4, one being the lower level of education. Despite the varying levels of education, the graph shows that the families remained around the same size. The median of each boxplot are somewhere between 2 and 4. What does differ, is the location of the third quadrant. It was more normal, or at least not considered an outlier for a woman's family with an education of 1 to have a family with more than 10 children then for any other level of education. The higher the level of education of the wife of the family, the less likely they were to have more than 10 children.

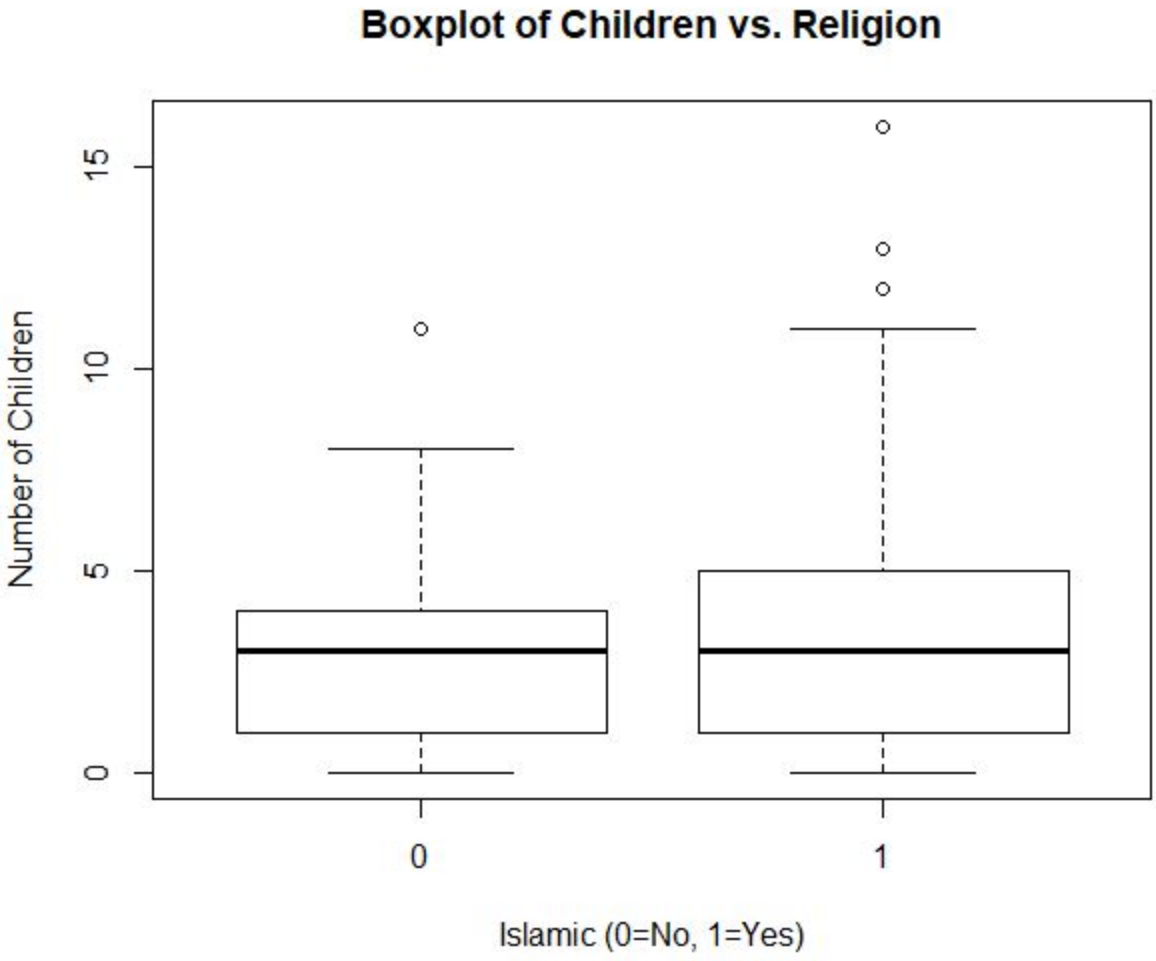


The husband of the family's level of education also had a similar effect on the number of children recorded. As the level of education of the husband increased, the median of the number of children in their family decreased. The third quartile of the boxplots also follow a similar pattern except for the 2nd ranking of education for the husband. Similar to what happened in the graph of the wife's education, what was an outlier for men with the fourth, and highest, ranking of education was not always an outlier for the first or second ranking of education. It is interesting to note also that the first and second levels of education for men do not even contain outliers as compared to men with higher levels of education.



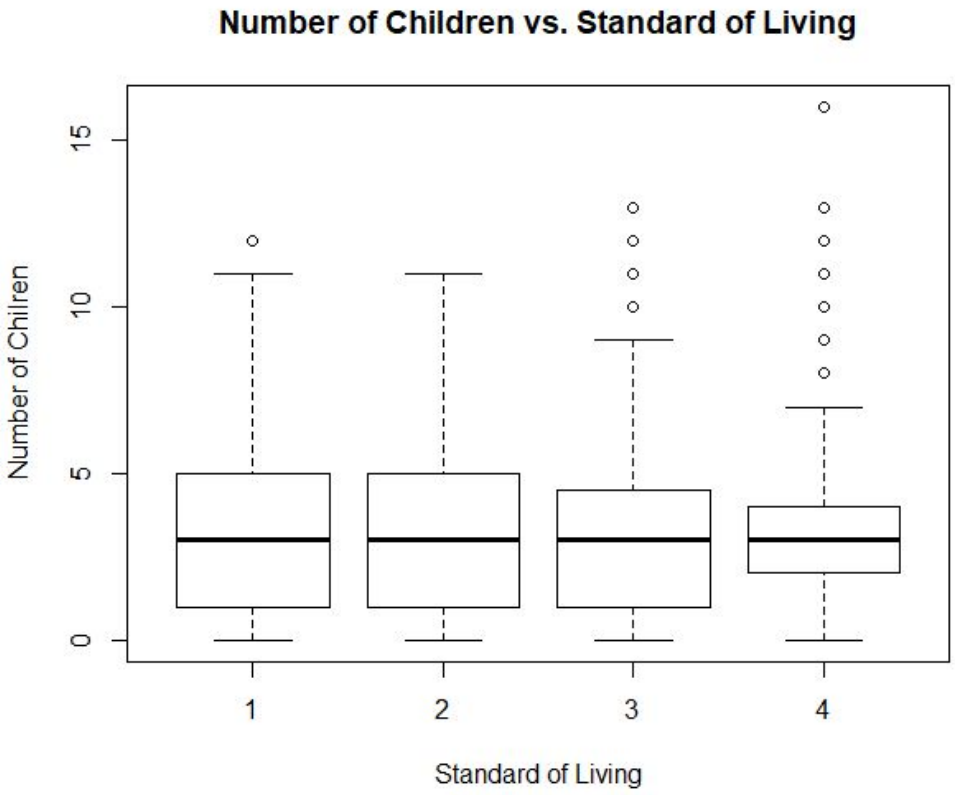
Another possible effect on the number of children in a family could be the religion of the family. Jokes are commonly made about catholic families being large compared to other families, but what is the effect of Islam on the size of families? Surprisingly, the median of the number of children for non-Islamic and Islamic families are the same. How if differs are the third

quartile and the cutoff line for outliers. The third quartile for non-Islamic families is around 4,

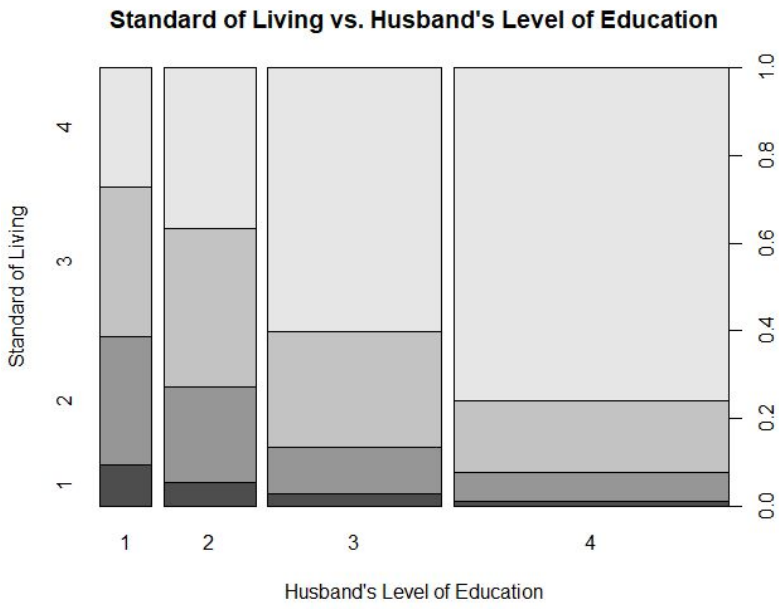


while for Islamic families is it 5. The major difference, however, is where the outliers lie. There is an outlier on the non-Islamic boxplot which would have been inside of the boundaries had the family identified as Islamic. Interestingly, the family who had more than 11 children were also all Islamic. So, even though the median of familial sizes does not change from one religion to another, it is more common to have larger families for people who identify as Islamic.

Another interesting effect on the number of children in a family is their standard of living. The ranking of the standard of living of different families goes from 1 to 4, 1 being the lowest and 4 being the highest standard of living.Though their medians may be similar, the cutoff for



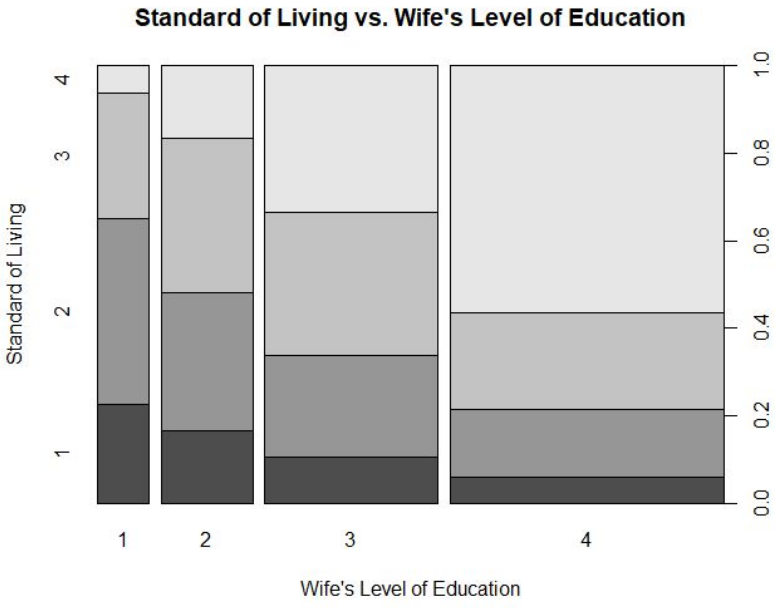
the outliers drastically changes from one standard to another, with the exception of rankings 1 and 2. As the standard of living increases from family to family, the cutoff for outliers decreases. As with the other graphs of the husband's education and the wife's education when the ranking increases there is a decrease in the number of families with large numbers of children.

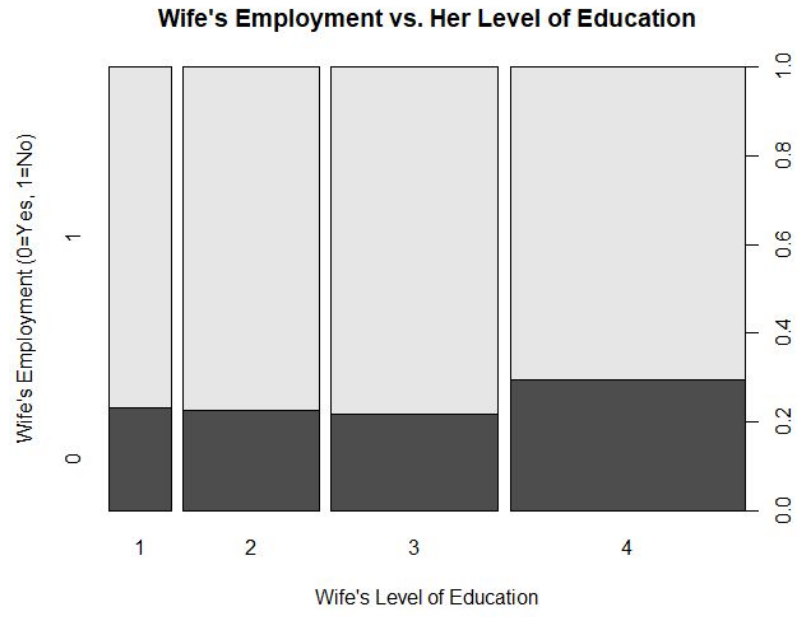


Another non-child related graph I thought would be interesting to look at is the standard of living compared to the husband's education. Does a higher education mean higher pay and a higher standard of living? As seen in the graph below, The majority of husbands with higher

level of education were able to provide a higher standard of living for their families. However this does not mean that only those with high levels of education could provide the highest standard of living. Over 20% of families with the lowest level of education were still able to provide the highest standard of living for their families and the percent only increases with the level of education of the husband. The correspondence between the family's standard of living and the husband's level of education was 0.357 by the spearman's correlation test.

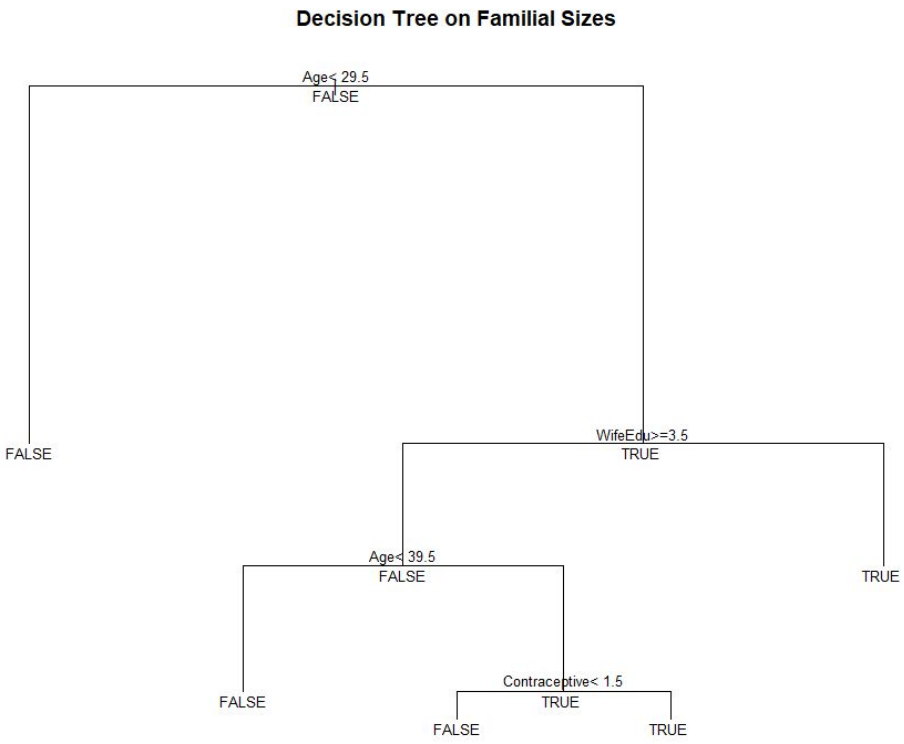
The graph is similar for the level of education of the wife of the family compared to their standard of living. As the wife's level of education increased, the percentage of families with higher standards of living increased. The spearman's correlation value for the family's standard of living compared to the mother's level of education was 0.3615. This is significantly higher that the value for the standard of living compared to the husband's education. Whether this is due to her having come from a wealthy family who could afford to provide that kind of education for their daughters we cannot determine with the data given. However we





can try to see if it is due to the wife working.

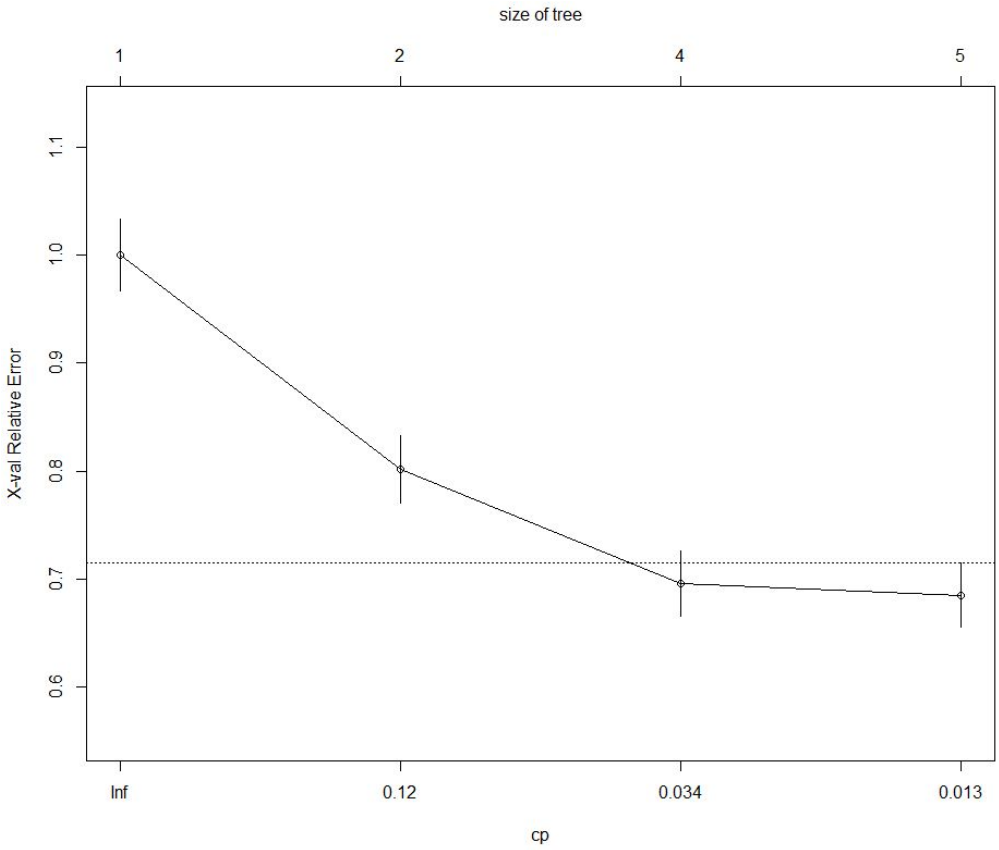
Below is a graph of the wife's level of education compared to whether she worked or not. There is no major change in the percentage of women who are employed compared to their



education except for between the third and fourth ranking of education. Despite there being change in the percentage of employment it is still minimal. The correlation between the wife's employment and her level of education is only -0.0622, significantly lower than the other two correlations observed. It is safe to conclude then that the wife's level of education does not affect her status of employment enough to create such a drastic change in the graph of the standard of living compared to the wife's education. Therefore the connection between the wife's level of education and standard of living must be from something else, or it is just an uncanny coincidence in the data.

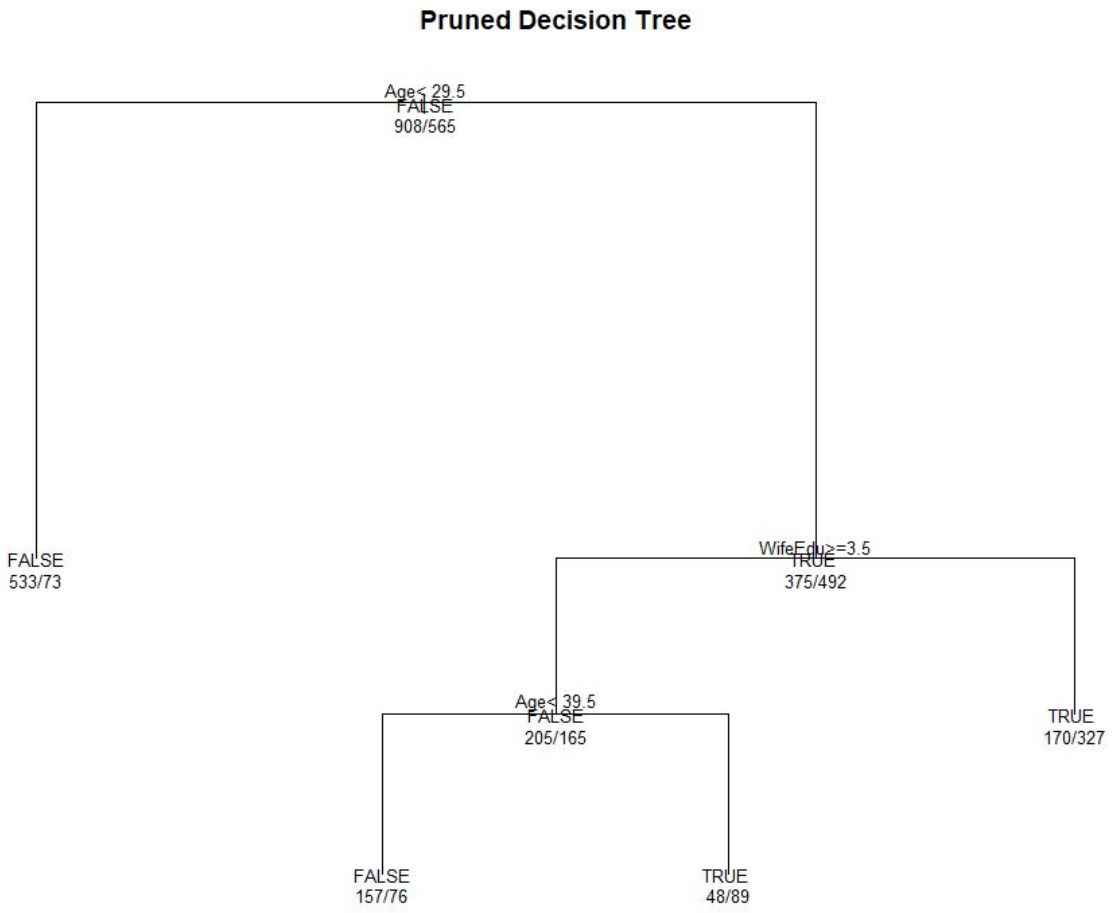
Next I decided it would be beneficial to try and look at a decision tree to determine what factors or combination of factors contributes to the likelihood of a larger family size. The decision tree would allow me to look at a few variables and see which ones were determined to be relevant. I chose to make any family with more than 3 children to be considered large since 3 was the median for the number of children in families as seen in the boxplot on pg 3. As seen in the graph below, the age of the mother, her level of education and the use of contraceptive for older women are the only factors that split the decision tree well. Despite having many factors, only these three matter. Pruning the tree is

Complexity Parameter Plot



also necessary and as seen from the graph below a good cutoff for pruning the tree would be somewhere after the first point that drops below the dotted line, but before the next point. I chose .02 as the complexity parameter to prune the tree.

The resulting decision tree had one less step than the one above. The pruned decision tree decided that the family's use of contraceptive was irrelevant, or that is did not help to split the data efficiently. Thus the only two factors that help split the data well are the wife's age and her level of education. As seen before in multiple plots, the likelihood that a family had many



children when the wife is young is very unlikely. When the mother is less than 30 years old,

there only 73 families recorded with more than the median of 3 children.The wife's education also assists in splitting the data well. For the variable of the wife's education, 4 is the highest. Therefore the second split in the decision tree is based off of whether or not the woman had the highest level of education possible or not. For those with lower than the highest level of education of the scale, the chance of them having an above-median family is larger. The split there is 66% in favor of having a larger family. For women who had the highest level of education, the left side of the split, only 45% of families had more than 3 children. The last split in the decision tree is done based off whether the woman was over or under the age of 39.5. For those aged 40 and over, the chances of them having a larger family was unlikely going all the way down to 32%. For the women aged under 40 and over 30 the likelihood that they have a larger family is significantly larger. The larger families made out 65% of the pool.

In conclusion only two variables really mattered when trying to determine whether a family's size would be over the median: the mother's age and her level of education. Even though there were correlations between a woman's level of education and their family's standard of living (0.3615 by the spearman test as stated above), the standard of living did not separate the family's probability of having a lot of children well enough. Women with lower than the best level of education and over the age of 29 had a higher likelihood of having a larger family. However if those women had the highest level of education (a 4 on the scale) and the women were over the age of 39 then they had a higher chance of having a larger than median family.

Certain effects on the data collected could be the kinds of people who decided to take the survey. Assuming that everyone told the truth on the survey, some people may not have had access to a computer or don't feel comfortable giving out their personal information. This could mean that data on the lower classes, those who cannot afford a computer, are not properly

represented on the data analyzed. Another factor that could make the analysis of the data less relevant is that the survey was performed in 1987, which was 30 years ago. Familial traditions, religion, levels of education and familial sizes all may have had a shift in trends in the past 30 years that are not represented in the data set. There was a dramatic decline in birth rate in the 1980s meaning that the size of families would not reach their peak as seen on graph 3 again. Here was yet another large drop in the country's birth rate in the 1990s, which, had the study expanded over various years, would have indicated a smaller size of families as time went on (3).

Works Cited:

(1) http://www.averagesalarysurvey.com/indonesia

(2) https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice

(3) http://countrystudies.us/indonesia/32.htm

Word Count: 2498