Hypothesis:

This project is divided into three main sections, each with a separate hypothesis statement:

Section1) There is a relationship between health care expenditure as % of GDP and infant mortality in developed and developing countries.
Section 2) There is a relationship between number of doctors and infant mortality in developed and developing countries.
Section 3) There is a relationship between number of doctors and infant mortality in the United States from 1991-2000.

Background:

The question of whether or not the gap between the developing and developed worlds is growing is one which has being increasingly researched and investigated in recent years. One such method of analysing this gap is through the study of certain health indicators, which can be used as representations of the overall well being of particular countries. This project is divided into three main sections, each incorporating different health indicators, and thus each section produces a different conclusion. The first section investigates the relationship between infant mortality rate and health expenditure in both developing and developed nations. Infant mortality rate is determined by calculating the number of infant (under 1 year old) deaths per 1,000 live births in a country, and health expenditure is given by the percentage of a country's GDP which is spent on health care. The second section investigates the relationship between the number of doctors and infant mortality rate in both developing and developed nations. The number of doctors in this section is a rate per 100,000 people, thus the number of doctors in a country for every 100,000 citizens. The third section specifically analyses the United States of America, and can be broken into two parts: Support ratio for public health care versus infant mortality; and number of doctors versus infant mortality.

Support ratio for public health care is determined by dividing the percentage of a state's spending on public welfare, health and hospitals, by the percentage of state's population with an annual household income below $15,000. Infant mortality in this section is a measure of the number of infant (below one year) deaths per 1,000 live births, and number of doctors is simply the number of registered medical doctors in the United States in a given year.

This project is based largely on comparing health indicators between *developing* and *developed* countries, two terms which are often referred to quite loosely. However, in this case, developing countries have been selected based on the definition of a "Least Developed Country (LDC)", which was created by the United Nations. This definition was created by the UN using: a low income criterion on GDP per capita; a human resource weakness criterion, based on health, nutrition, education, and adult literacy; and an economic vulnerability criterion, based on the instability of aspects of the economy, and merchandise export concentration.[1]

Procedure and Use of Technology:

All of the statistics used for this project were obtained from the internet, however all data was obtained from trustworthy, reliable sources, thus it can be considered highly accurate and dependable. The infant mortality rates for Sections one and two are from the World FactBook, a yearly publication of pertinent data on every world nation. Also for sections one and two, the health expenditure and number of doctors are from a very reliable source, the World Health Organization, which is a United Nations specialized agency for health. For section three, infant mortality data was obtained from the National Center for Health Statistics and the United Health Foundation, while expenditure and number of doctors was from the Center for Justice and Democracy. Though these three sites may not be as highly regarded as the World Health Organization (WHO), the data obtained was compared with, and is consistent with similar such data from the WHO and World FactBook, thus validating its reliability. Statistics were found quite easily on the WHO and World FactBook websites, for they are very well organized, and provided

extensive statistics. However, finding the specific data for the United States was more difficult, and health personnel for individual states and number of doctors over various years within the USA could not be found on the WHO or World FactBook sites. Therefore it was necessary to use an internet search engine, and open the sites to all search results until appropriate data was found from a seemingly reliable source.

Technology was also used in creating graphs and performing various calculations used to answer the posed hypothesis questions. Microsoft Excel was the preferred method of technology, and through this program all of the raw data was organized, and scatter graphs and bar graphs were created. Furthermore, Excel was used to calculate the mean values, as well as the coefficients of determination (r^2) and correlations coefficients (r) for each scatter graph. Once data was found from a website, it was entered into a specific Excel spreadsheet (depending on the hypothesis question and section to which the data pertained) from which calculations could be performed and graphs constructed.

As well as Microsoft Excel, Fathom was a program used to analyze the data sets specific to only parts of Section 1 and Section 3. For Section 3, support ratios for each state and infant mortality for each state were plotted as box and whisker plots to determine which states were potential outliers to each factor. Then, support ratio and infant mortality were plotted as a scatter graph, where a linear regression could be conducted and an r^2 value calculated. Fathom was very easy to use for this process, as when I wished to determine which state was a particular point, and the values of each point, the point could simply be clicked on, and the values would be highlighted in the collection. Box and whisker plots were also created for Section 1, to determine if there were outliers for both health expenditures and infant mortality in developing countries. These box plots are effective as it is clear just from looking at the graph if an outlier to that factor exists, and it is simple and fast to use Fathom for this type of graph.

Calculations:

Explanation of Calculations:

1) Correlation Coefficient (r) =

This is a measure of how well a line fits a particular set of data, or the strength of a linear correlation. It is calculated by dividing the covariance by the product of the standard deviation of X and the standard deviation of Y. (Covariance is the product of the deviations of X and Y divided by one less then number of data points.) Thus the correlation coefficient measures how closely data points cluster around a line of best bit. The r value, also called the Pearson product-moment coefficient of correlation (Pearson's r), ranges from –1 to 1, and the closer the value to 1, the better the fit. If $-0.33 < r < 0.33$, the correlation is weak, if $-0.67 < r < -0.33$, or $0.33 < r < 0.67$, the correlation is moderate, if $r < -0.67$ or $r > 0.67$, the correlation is strong, and if $r=0$, there is no correlation between the two variables.

2) Coefficient of Determination (r^2) =

This is a measure of how closely a curve fits a particular set of data, and is calculated by measuring the distance of the residual data points that do not fall directly on the line of fit. The distance of all residual points are squared and summed to determine the r^2 value. This measure explains the relationship between the amount of variation in the response variable due to the variation in the explanatory variable. Values fall between 0 and 1, and the higher the value, the better the fit of the curve. Though this value can be used as a measure for all types of regression, in this particular project all regressions are linear, thus all r^2 values represent a linear line of fit.

Presentation of Data:

Section 1: Health Expenditure versus Infant Mortality, Developed and Developing Countries

a) Health Expenditure versus Infant Mortality, Developing Countries

Correlation Coefficient (r): -0.382
Coefficient of Determination (r^2): 0.146

The r value of these variables show that there is a moderate negative correlation between health expenditure and infant mortality in developing countries, however this value is on the weak side of moderate, as it is very close to 0.33, which is the boundary value between weak and moderate correlation according to Pearson's r. The r^2 value also supports this weak/moderate correlation, as it is only a value of 0.146. Thus, only 14.6% of variation of the infant mortality rate is due to the variation in health expenditure.

b) Health Expenditure versus Infant Mortality, Developed Countries

Correlation Coefficient (r): 0.559
Coefficient of Determination (r^2): 0.312

In developed countries, there is a moderate positive correlation between health expenditure and infant mortality according to the ranking of the r value (0.559) by Pearson's r. As well, the r^2 value is 0.312, meaning that 31.2% of the variation in infant mortality in developed countries is due to the variation in health expenditure. This value differs from part a, as the correlation is positive, indicating that as health expenditure increases, infant mortality increases as well.

c) Health Expenditure versus Infant Mortality, Developed and Developing Countries

Correlation Coefficient (r): -0.837
Coefficient of Determination (r^2): 0.701

The combination of data from developing and developed countries results in a strong negative correlation between health expenditure and infant mortality. The r value is –0.837, which according to Pearson's r indicates a strong relationship between the two variables. Also, the r^2 value is quite high, at 0.701, meaning that 70.1% of the change in infant mortality is explained by the change in health expenditure. Of this section, the combination of developing and developed countries shows the strongest correlation of health expenditure and life expectancy.

Section 2: Number of Doctors versus Infant Mortality, Developed and Developing Countries

a) Number of Doctors versus Infant Mortality, Developing Countries

Correlation Coefficient (r): -0.347
Coefficient of Determination (r^2): 0.121

In developing countries, there is a moderate negative relationship between number of health personnel and infant mortality. This number is so close to -0.33 however, that the relationship is very close to being weak. Also, the r^2 value is quite low (0.121), meaning that 12% of the change in the infant mortality rates in developing countries is explained by the change in the number of health personnel.

b) Number of Doctors versus Infant Mortality, Developed Countries

Correlation Coefficient (r): 0.0954
Coefficient of Determination (r^2): 0.0091

In developed countries, there is a *very* weak negative relationship between number of doctors and infant mortality. This is justified by the r value of only 0.0954. This value is very close to zero, which would indicate *no* correlation at all between the two variables. Also, the r2 value is 0.0091, meaning that only 0.91% of the change in infant mortality is explained by the change in number of doctors. These values are so low that there is virtually no correlation between number of health personnel and infant mortality in the developed world.

c) Number of Doctors versus Infant Mortality, Developing and Developed Countries

Correlation Coefficient (r): -0.88
Coefficient of Determination (r^2): 0.774

The combination of data from developing and developed countries indicates a strong negative correlation between number of doctors and infant mortality. The r value of -0.88 is, according to Pearson's r, in the range representing a strong correlation, and the r^2 value is high, at 0.774, meaning that 77.4% of the change in infant mortality is due to the change in number of health personnel.

Section 3: Support Ratio and Number of Doctors versus Infant Mortality, United States

a) Support Ratio for Public Healthcare versus Infant Mortality

Correlation Coefficient (r): -0.383
Coefficient of Determination (r^2): 0.1469

In the United States, there is a moderate negative correlation between a states support ratio for public health care and infant mortality rate, as the correlation coefficient representing all 50 states is –0.38. Furthermore, the r^2 value for this data is only 0.1469,

meaning that only approximately 14.7% of the infant mortality rate is explained by the support ratio for public healthcare.

b) Number of Doctors versus Infant Mortality

Correlation Coefficient (r): -0.971
Coefficient of Determination (r^2): 0.9437

The negative correlation between the number of medical doctors and infant mortality in the United States from 1991-2000 is much stronger than the correlation between part a, support ratio for public healthcare and infant mortality. This correlation is very strongly negative, and almost perfect according to Pearson's ranking, as the r value is –0.97. As well, the r^2 value is very high (0.94), meaning that 94% of the infant mortality in the United States is explained by the number of medical doctors. Furthermore, this data contains no outliers, which contributes to the strength of this linear regression.

Conclusions:

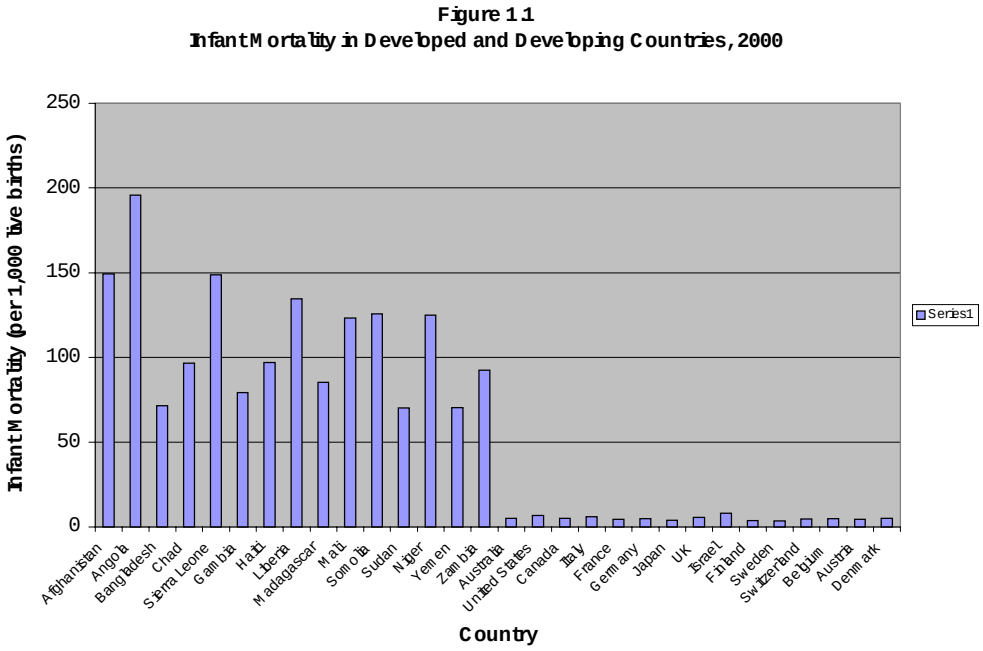
Section 1- Health Expenditure and Infant Mortality, Developed and Developing Countries

Prediction:

I predicted that for both developing and developed countries there would be a strong negative correlation between health expenditure and infant mortality; therefore as health expenditure decreased, infant mortality would increase. Also, I predicted that there would be a large gap between the money spent on health care between the two worlds, as well as a large gap between their infant mortality rates, with the developing nations having higher health expenditure and lower infant mortality rates.

Conclusion:

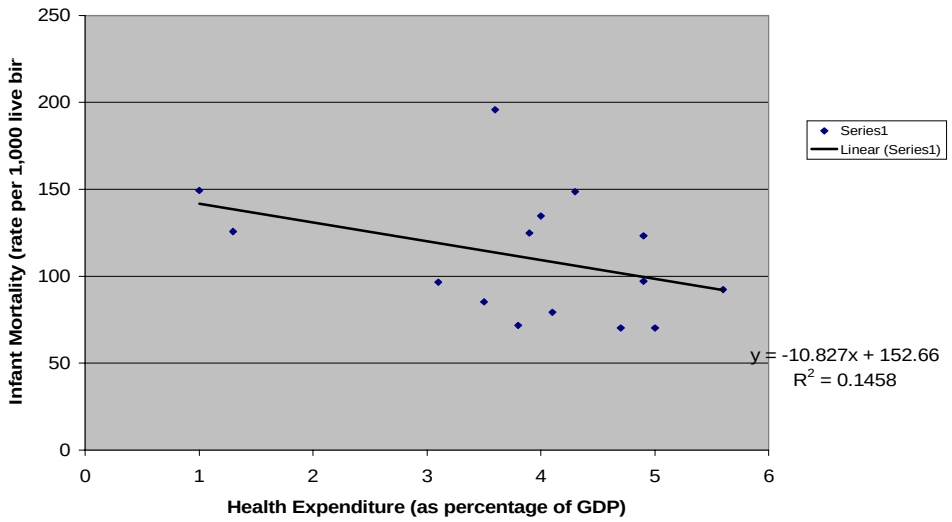
My prediction that there would be a large gap between the health expenditures and infant mortalities in developing and developed nations was certainly proven by Figure 1.1.



This figure, representing health expenditure as a percentage of GDP shows that in developing countries the percent of GDP spent on health care ranges from 1%-5.6%, while in developed countries, the percent of GDP spent ranges from 6.6% to 13%. The mean percent spent on healthcare for developing countries is 3.85%, and 9.02% for developed countries. Furthermore, the infant mortality rates also vary greatly, as the rates in developing countries range form 70.21-195.78 (with a mean of 111), which is significantly higher than the range in developed countries, which is 4.51-7.9 (with a mean of 5.1).

Though I was accurate in predicting the large gaps in health expenditure and infant mortality rates in the two worlds, my predictions on the correlations were not as accurate.

Figure 1.2
Health Expenditure versus Infant Mortality in Developing Countries, 2000



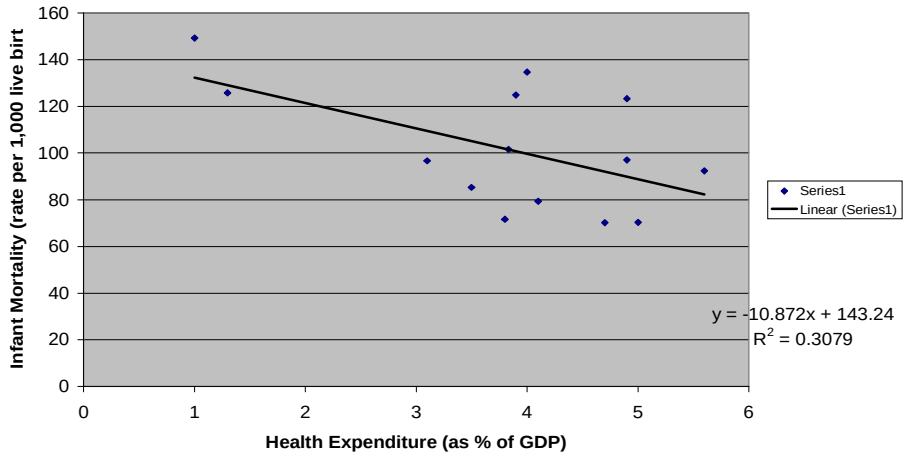
The correlation between these two variables for developing countries (Figure 1.2) shows a weak to moderate negative correlation. It is understandable that this relationship is negative rather then positive, since it is logical that as a country contributes a larger percentage of its GDP to its healthcare system, it is able to provide better services for its citizens. One primary service that this money goes towards is hospitals and health care centers. It is in these centers that women can receive check-ups during pregnancy and be forewarned of in illness in a child that they are carrying so that treatment can be provided even before a child is born. Furthermore, when a baby is born in a hospital under proper care, they are able to receive proper treatment following birth.

It is surprising, however, that the r value for this correlation is so low, at only -0.382. A possible explanation for this low value is the presence of outliers in the data set. One such outlier is Angola, which has a health expenditure value almost identical to the mean value of 3.84, but also has an infant mortality rate far higher then any of the other developing countries. Thus this particular data point falls far above the line of best fit, and skews the data. The removal of this particular point raises the correlation coefficient value from –0.381 to –0.455, thus resulting in a stronger relationship between the two factors. A r value of -0.455 means that there is a moderate negative relationship, as opposed to the previous value which was on the borderline of indicating a weak negative relationship. Angola's high infant mortality can be justified however, as it a

country plagued by many of the main causes of infant mortality (according to the World Health Organization), including pneumonia, malnutrition, HIV/AIDS, malaria, measles and diarrhoea.[2] Also, there is little nutritious food available in Angola for mothers, meaning in turn the breast milk used to feed children is not sufficiently nutritious. Especially in early months of life, nutritious breast milk is crucial, and a lack of this can help contribute to high infant mortality. Also, if a mother with AIDS gives birth, there is a great chance that her child is born with the virus as well. Furthermore, Angola is a country in which 20 years of recent war have ruined 70% of its health infrastructure and sanitary network, meaning few services are available for mothers and their children.[3]

Another outlier which does not have as significant of an effect as Angola is Sierra Leone. Sierra Leone has a justifiably high infant mortality also for reasons similar to those affecting Angola. It's citizens also experience high levels of fever, malaria, diarrhoeal disease, respiratory infections and HIV/AIDS, however perhaps the greatest factor is the civil war in Sierra Leone.[4] It has been in conflict since 1991, and though the UN has attempted repeatedly to intervene, much of this has failed to result in any solutions. Civil conflict in a country tends to increase not only infant mortality, but overall mortality rates. Also, medical care in Sierra Leone is out of date and expensive, and as well there is a great shortage of drugs and equipment. All of these factors may contribute to the high infant mortality rate in Sierra Leone.[5]

Figure 1.3
Health Expenditure versus Infant Mortality in Developing Countries, Removal of Angola and Sierra Leone



When Sierra Leone is removed along with Angola, shown in Figure 1.3, there is a considerable change in the strength of the correlation between the two variables. The correlation coefficient changes from -0.382 to -0.55, and the r^2 value changes from 0.146 to 0.308. Therefore this relationship changes from being almost weak to clearly moderate. Thus these two countries are outliers, as their removal creates a correlation which appears to be a more accurate depiction of this relationship.

Another way in which outliers can be determined is through the plotting of variables in box and whisker plots, which when using Fathom, clearly indicate outliers as dots not included on one of the four quartiles. Thus, both health expenditure and infant mortality were plotted as Figures 1.4 and 1.5 respectively, in box and whisker plots.

Figure 1.4

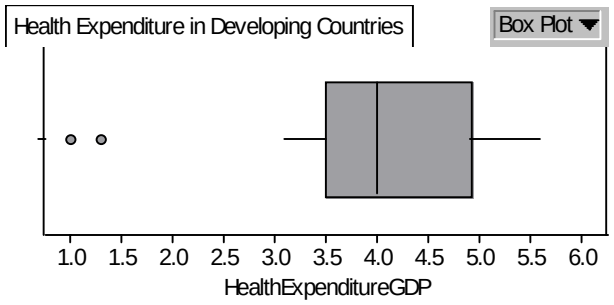
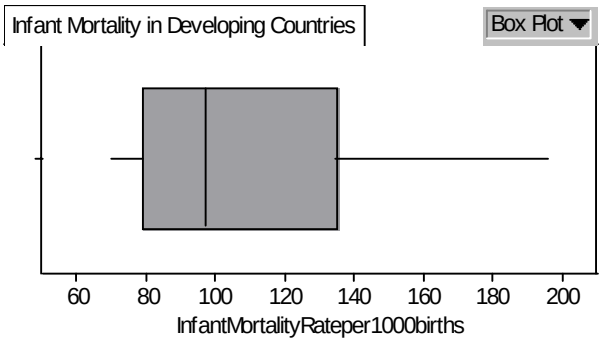
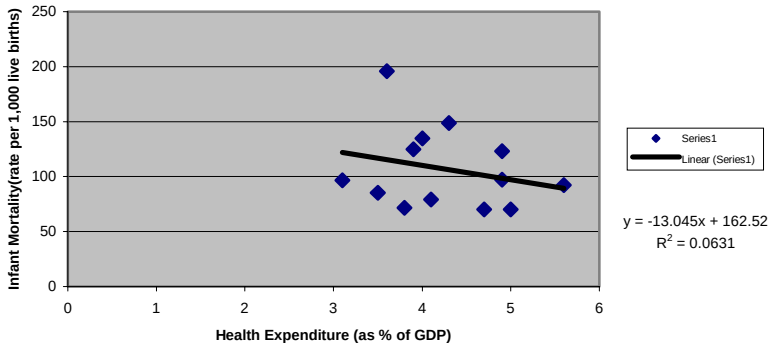


Figure 1.5



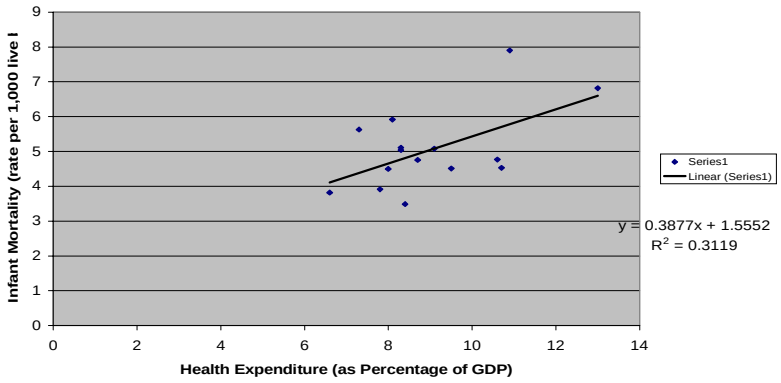
Though Figure 1.5, showing infant mortality does not indicate any outlying points, Figure 1.4 indicates that there are two outliers to the data, which are Afghanistan and Somalia. These points are outliers for they have very low relative health expenditures: 1% in Afghanistan and 1.3% in Somalia. These health expenditures are well below the third lowest point, which is 3.1%, thus validating that they are outliers.

Figure 1.6
Health Expenditure versus Infant Mortality, Developing Countries, Removal of Afghanistan and Somalia



However, the removal of these two points shown in Figure 1.6 indicates that they are not outliers that weaken the strength of this correlation. Their removal decreases the correlation coefficient from –0.382 to –0.25, and the r^2 value from 0.146 to 0.0631, therefore they actually contribute to the strength of this relationship, not the weakness.

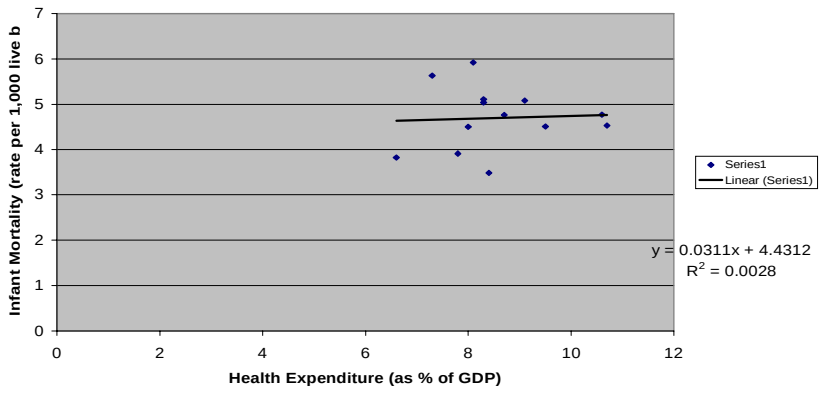
Figure 1.7
Health Expenditure versus Infant Mortaliy in Developed Countries, 2000



The regression conducted on the 15 chosen developed countries (shown in Figure 1.7) produced a result quite opposite to that of the developing countries, as initially the correlation in developed countries was positive, not negative as predicted and expected. In fact, this positive correlation has an r value of 0.56, indicating a moderately strong relationship. This result seems to make little sense (especially with the results in developing countries), as it is logical that as a country contributes more money towards its healthcare, its infant mortality rate would decrease, not oppositely as shown by this result.

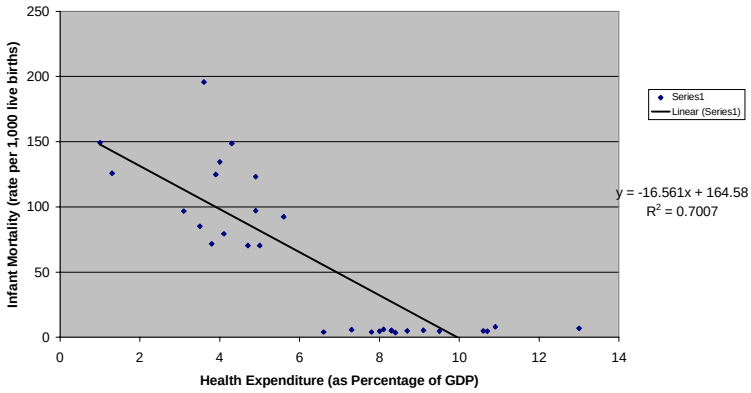
However, this theory is disproved both with the initial regression, and the regression performed after the removal of 2 outliers. Israel and the United States are considered outliers, as on the scatter plot they are points that are quite far from any of the other 13 points. The United States has the second highest infant mortality rate (next to Israel), but also has the greatest health care expenditure. Though the USA is an extremely rich country that contributes greatly to its health care, there is also a large disparity in the distribution of wealth, and it is this disparity that creates certain rich areas, but also many impoverished areas with little healthcare. It is in these impoverished areas that the infant mortality would be high, and outweigh the infant mortalities in the richer areas. Another outlier, Israel, has the highest infant mortality rate and contributes the second most to healthcare (next to the USA). Though Israel is a developed country, it is also plagued by war (and had been for decades), thus increasing its overall mortality as well as its infant mortality.

Figure 1.8
Health Expenditure versus Infant Mortality, Developed Countries, Removal of Israel and United States



Analyzing the graph along with Israel and the USA would indicate that there is a positive relationship between these two factors in developed countries. However, when they are removed from the data set (Figure 1.8), the correlation does not become negative as in the developing countries. Instead, the correlation coefficient value drops from 0.56 to 0.05, indicating that there is *no* relationship between these two factors, since 0.05 is so close to zero. Furthermore, the r^2 value before the outliers were removed is 0.3, and drops also to 0.0028. This number is so small that virtually none of the change in infant mortality rate can be explained by the change in health expenditure.

Figure 1.9
Health Expenditure versus Infant Mortality in Developing and Developed Countries, 2000



However, though there appears to be large differences between this correlation in developing countries and developed countries, it is the combination of the two, shown in Figure 1.9 that indicates the overall world-wide relationship. This correlation results in the strongest relationship between health expenditure and infant mortality, with a correlation coefficient of –0.84 and an r^2 value of 0.7. As well, this result solidifies the conclusion that this is a negative relationship, not positive as found in developed countries. Perhaps there was large discrepancy between the 3 data sets pertaining to this question simply due to the countries that were chosen for this study. There are more than 15 developing countries and more than 15 developed countries, however not every country in the world was used for this project. More accurate results would be achieved if every country was used, as it is very easy for data to be skewed, and inaccurate conclusions to be drawn when only 15 countries are used for each graph. As shown in this project, one country can have a great effect on conclusions, and the more countries you use, the less of an effect an outlier will have.

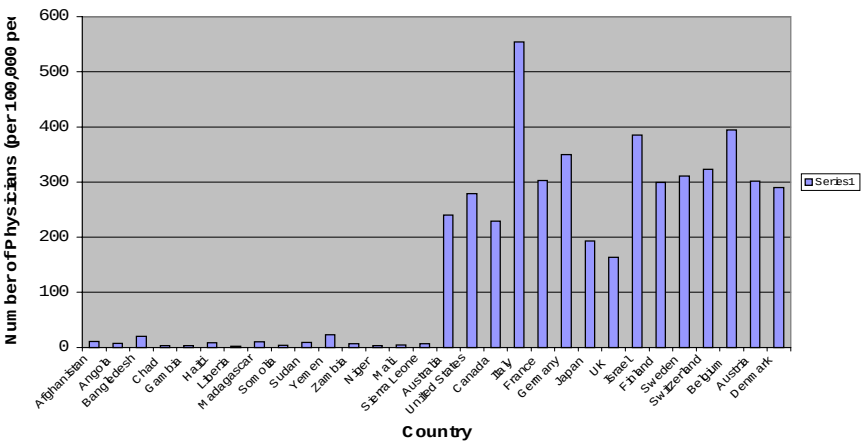
Section 2- Number of Doctors and Infant Mortality, Developed and Developing Countries

Prediction:

Similar to section one, I predicted that the correlation between number of doctors and infant mortality in developing and developed countries would be strongly negative, therefore as number of doctors increases, infant mortality would decrease. Also I predicted there to be a large gap between the developing worlds with regards to both number of doctors and infant mortality, with the developed nations having more doctors and lower infant mortalities.

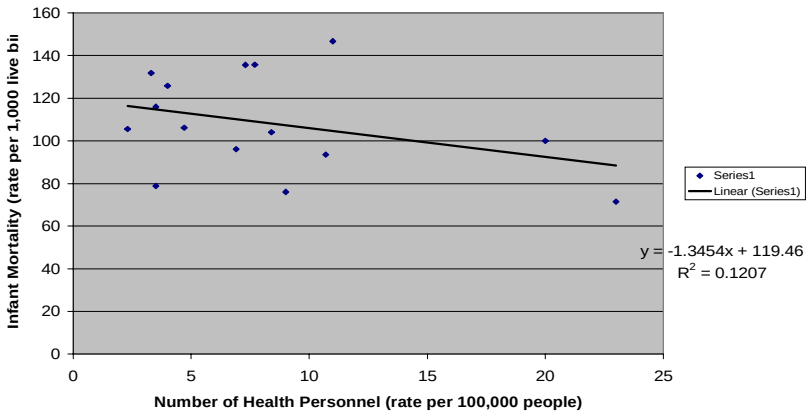
Conclusion:

Figure 2.1
Number of Doctors, Developed and Developing Coountries



It is clear from looking at Figure 2.1, that there is a very large difference in the number of doctors in developing and developed countries. This can be explained also in combination with the heath expenditure data from Section one, as countries spending less on health care also have fewer doctors, which makes sense. A country spending less on healthcare would have fewer hospitals and/or clinics, and it would be in these locations that doctors would be required. In the developing countries, number of doctors per 100,000 people ranges from 2.3-23, while in developed countries this range is from 164-554, which is greatly higher.

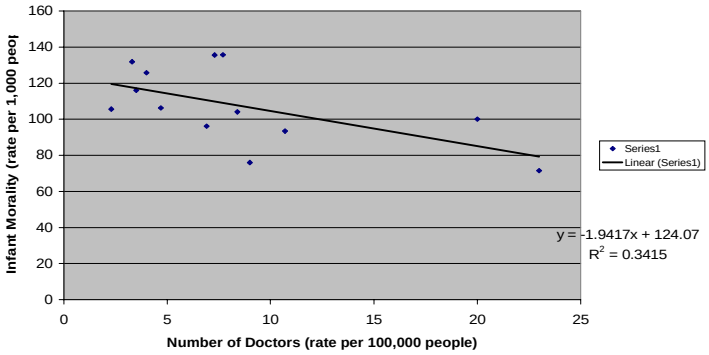
Figure 2.2
Number of Doctors versus Infant Mortality, Developing Countries



Similar to Section one, I predicted a very strong correlation between number of doctors and infant mortality in developing countries, however this was not the conclusion that was reached from analysing the data. The correlation coefficient for developing countries was –0.347, a value very close to the corresponding value for health expenditure. Certain countries are also clear outliers in Figure 2.2, and are plotted fairly far from the line of best fit. The two most dramatic outliers are Afghanistan and Gambia, as Gambia falls far below the line, and Afghanistan falls far above the line. Gambia has one of the lowest doctor to population ratios, yet it also has one of the lowest life expectancies of the developing countries studied, which indicates a positive relationship. Gambia's low infant mortality rate can be explained by its well established healthcare centers which are located not only in more urban areas, but in remote villages and communities. Also, nurses and birth attendants are stationed in more local villages to aid expected mothers in their pregnancies. This system in Gambia is unlike any other African country, and decreased their infant mortality by one half from 1982-1994. People in Gambia are learning the value of healthcare, and this is certainly having a positive effect on its infant mortality.[6] Oppositely, Afghanistan has more doctors per 100,000 then the mean, yet has a life expectancy far above the mean, and far above any of the other developing countries, which also indicates a positive relationship. Afghanistan is a country which has recently experienced war, but moreover, it is a country plagued with poor health. In 2002 it had a 58% prevalence of chronic malnutrition among children under five, and other major causes of mortality include

diarrhoea, respiratory infections, and malaria.[7] Furthermore, women in Afghanistan tend to have low Iodine counts, causing low birth weight in children, which leads to more frequent infant deaths.[8]

Figure 2.3
Number of Doctors versus Infant Mortality, Developing Countries, Removal of Gambia and Afghanistan



The removal of Gambia and Afghanistan significantly increase the strength of this relationship, as the correlation coefficient changed from –0.347 to –0.58. Before their removal, the relationship was only weakly to moderately negative, but after their removal, the relationship became clearly moderate, and even close to strongly negative. Also, the r^2 value changes from 0.121 to 0.34, which shows an increased strength in the regression line. This is almost identical to the relationship between health expenditure and infant mortality, as both started as weak/moderate correlations, and increased to moderate when outliers were removed. This similarity is one way to validate the results of this project, as both health expenditure and number of doctors basically affect infant mortality in the same way for developing countries.

Figure 2.4
Number of Doctors versus Infant Mortality, Developed Countries

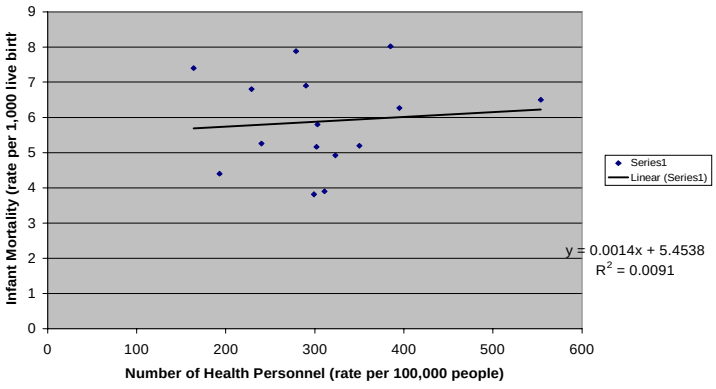


Figure 2.4, representing number of doctors versus infant mortality in the 15 developed countries, shows a weak positive relationship between these two variables. However, the correlation coefficient of 0.0954, and coefficient of determination of 0.0091 are so low, that there is virtually no correlation between the number of doctors and infant mortality. Furthermore, there are no significant outliers to this graph, as all of the point are quite scattered around the line of fit, and also quite far away from the line. This result is also similar to the corresponding graph for health expenditure in developed countries, thus perhaps validating the accuracy of this correlation. In developed countries, economies are so varied, and there are so many extraneous variables that it is virtually impossible to conjure a strong relationship.

Figure 2.5
Number of Doctors versus Infant Mortality, Developing and Developed Countries

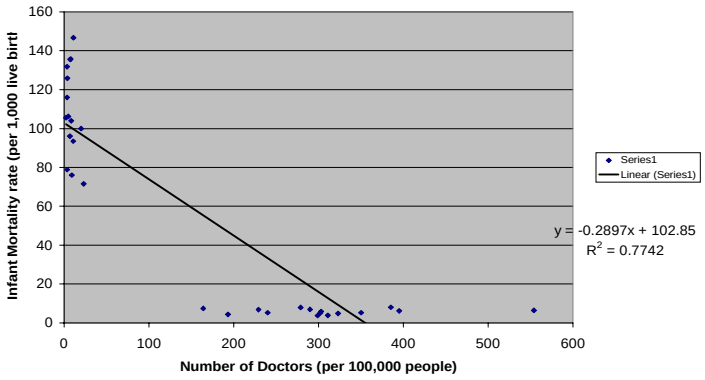


Figure 2.5, representing number of doctors versus infant mortality for developing and developed countries illustrates a strong negative relationship between these two variables, with an r value of -0.88 and an r^2 value of 0.774. These results are justifiable, as there is such a gap in the data between the developed and developing worlds, that it makes sense for this data to be strongly negative. Data points on the graph appear at two different extremes, with none in the middle. Overall, this graph presents an accurate overall indication of the health disparity in the world, and also proves the negative relationship existing between number of doctors and the infant mortality.

Section 3- United States Support Ratio for Public Health Care versus Infant Mortality, and Number of Doctors versus Infant Mortality, United States

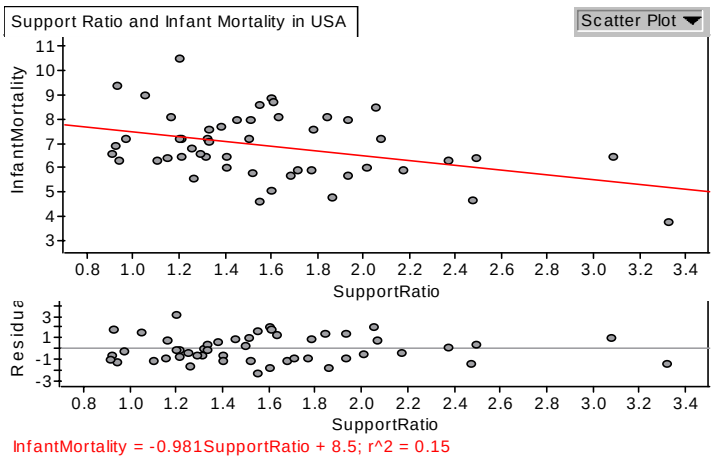
Prediction:

For this section I did not think that the correlation between support ratio and infant mortality would be as strong as the correlation between health expenditure, health personnel, and infant mortality for various developing and developed countries in Sections one and two. Though I was aware of the large wealth gap within the United States, I did not believe that a state's contribution to health care would ultimately have a large effect on infant mortality, due to the extensive privatization of the health care system. Thus I believed that only to a moderate extent would infant mortality increase as support for public health care increased. I also thought that the change in the number of medical doctors in the United States would have a direct effect on infant mortality, for it seemed very obvious to me before even collecting data that the number of doctors is steadily increasing (also as the American population increases), while infant mortality is decreasing.

Conclusion:

Part A: Support Ratio versus Infant Mortality

Figure 3.1

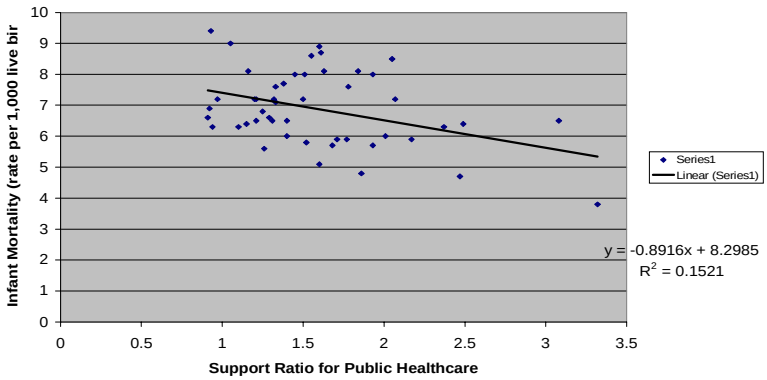


Though I did not necessarily expect the correlation between these two factors to be strongly negative, I did not think that the correlation would be this weak. However, upon further analysis of the raw data and scatter graph, this weak correlation appears to make more sense, due to the effect of outliers. When simply looking at the scatter plot of these two factors (Figure 3.1), it seems that the graph has two clear outliers, which would significantly skew the data. One outlier seems to be Maine, with a support ratio of 1.55 and an infant mortality rate of 4.6. This state would appear to skew the data because though its support ratio is virtually the same as the mean value of 1.59, its infant mortality is 2.3 deaths below the mean value of 6.9. Maine in fact has the lowest infant mortality rate of any state, even though its support ratio is approximately average. This reflects the fact that Maine is a state with an environment that "nurtures pregnant women and their unborn children."[9] Thus on the scatter graph, the point is far below the line of best fit, which ultimately pulls the line downwards, and may decrease the correlation between the two variables.

A second apparent outlier is Mississippi, which clearly has the highest infant mortality rate of all 50 states, at 10.5. This value is 3.6 deaths above the mean value of 6.9. This point is an outlier however because though it has the highest infant mortality rate, is does not have the lowest support ratio of all states, as eight other states have support ratios lower than that of Mississippi (1.2). This high infant mortality is largely due to the problem that Mississippi faces with both poverty and high rates of teenage pregnancy. Facing these conditions, many women are not adequately prepared to deal with the responsibility of a child, and may not be able to sufficiently support it. A lack of proper family structures also contributes to this as it is a family support system that nurtures mother and babies, and provides love and aid where necessary. With this element missing, babies certainly have less of a chance of living.[10]

If this data set created a strong correlation, Mississippi would not only have the highest infant mortality rate, but it would also have the lowest support ratio for public healthcare.

Figure 3.2
Support Ratio for Public Healthcare and Infant Morality, Removal of Maine and Mississippi



However, these two states are actually not outliers, because when the correlation coefficient was recalculated with their removal (Figure 3.2), it increased by only 0.006, which essentially is no change at all. Therefore, Maine and Mississippi are points which do not skew the data, as they're removal does not strengthen the correlation between support ratio and infant mortality.

Figure 3.3

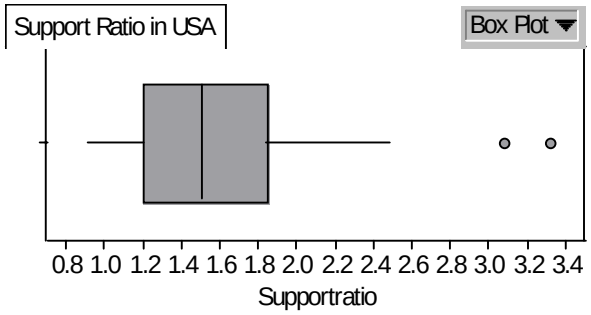
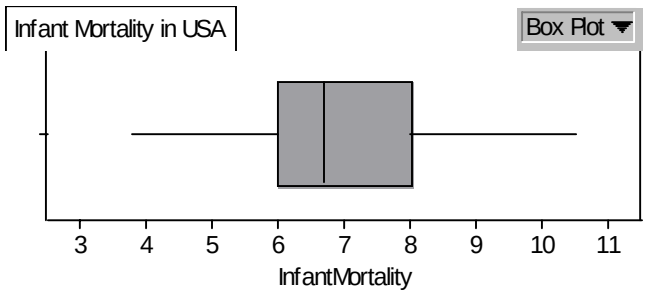
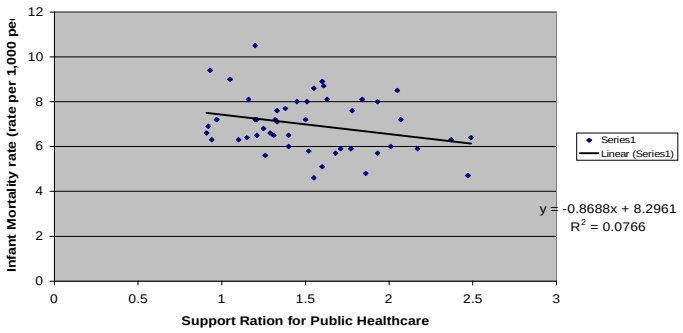


Figure 3.4



Another method used to determine outliers which may have skewed this data is through figures 3.3 and 3.4, which are box and whisker plots of support ratio and infant mortality (respectively) for all 50 states. Figure 3.3 showing support ratio, indicates that there are two states which are outliers to the data, Colorado and New Hampshire. Both of these states have support ratios which are very high in comparison to the other states. Figure 3.4, showing infant mortality, does not indicate any outliers to this factor.

Figure 3.5
Support Ratio versus Infant Mortality in USA, Removal of New Hampshire and Colorado



After completing the linear regression again with the removal only of Colorado and New Hampshire, it became clear that these two states are not outliers, as their removal decreased the strength of the negative relationship. The r^2 value decreased from 0.147 to 0.077, meaning the change in infant mortality became explained even less than the change in infant mortality. Also, the r value changed from –0.38 to –0.277, which is closer to zero, or a non-correlation between variables. Their removal did not make this relationship more strongly negative as one may have predicted, thus they do not act at outliers which would skew the data set.

Part B: Number of Doctors versus Infant Mortality, 1991-2000

Figure 3.7 indicates a very strong negative relationship between these two variables, with a r value of -0.971 and an r^2 value of 0.944. These results are not particularly surprising, for the number of doctors from 1991-2000 has increased at a steady rate, while the infant mortality has decreased at a steady rate. In every year that the number of doctors has increased, the infant mortality has decreased; and the situation never occurred where the number of doctors increased as well as infant mortality. As the number of doctors increases in the United States, it makes sense that the country would see benefits in their healthcare system. More doctors primarily means that more attention can be paid to the monitoring of infant health in hospitals and within the first few vital months of their lives. Beyond this however, in a developed country such as the United States, more doctors means that more research can be conducted, and more cures or drugs for diseases can be developed which could prevent early death of infants. Such developments could all contribute to a steadily lowering infant mortality rate over time. This relationship is stronger than any other in this project perhaps because it is dealing within one particular country (as opposed to 14 other countries), and also not dealing with 50 different states. Analyzing a variety of locations is difficult because there are so many extraneous variables pertaining to each particular location, and the elimination of these variables is virtually impossible. However, this correlation is very simple, and is covering *only* the United States, thus increasing the chances of creating a strong correlation.

Assumptions:

The greatest assumption in this project is that all data collected is accurate, and therefore that the conclusions reached are relevant. This potential inaccuracy of statistics applies especially to the infant mortality rates in developing countries. In developing nations, the great problem is that few births of babies are actually reported or registered, which is greatly due to the small number of hospitals. Births are mainly reported only when a baby is born in a hospital, however for many babies, this is not the case. Thus, if many births are not reported/registered, then the deaths of babies either at birth or within the first year of life are often not reported either. This would lead to an infant mortality rate that is not an accurate representation of the number of infant deaths in the country. If many deaths are not reported, then the infant mortality rate found on the WHO website may actually be too low.

Another assumption pertaining to infant mortality rates is that countries are not involved in war, being a civil war or a war with another country. An countries involvement in war would ultimately lower the infant mortality rate, thus contributing to the inaccuracy of statistics.

Limitations:

One limitation of this project which proved to pose many problems and frequently interfere with my progress was that many of the data sets I found throughout my investigation were incomplete, and for particular countries did not provide necessary data. One such example of this problem was with the World Health Organization data on Health Personnel, as for many of the small, developing nations there were no figures provided for the number of health personnel. Considering the few number of developing countries I could possibly study in the first place, this severely lowered this number. Also since I completed this stage of the project *after* I had analyzed health expenditure, I had to go back and change a number of the developing countries which I had chosen for the health expenditure. I did this because I wanted to maintain consistency from section

to section of this project, and one way to maintain consistency is by analyzing the same countries for each section. To solve this problem, I went back to my first section on health expenditure and changed the developing countries I used, making sure that I had complete data for each of these countries.

Another limitation of the data used applies mainly to the developing nations used, which is the accuracy of the data provided. Although for Sections one and two, very reliable sources were used, it is still likely that the figures were slightly off. According to Washington University, infant mortality rate is one of the "most incompletely recorded events", though it is one of the most significant and frequently studied health indicators. Infant mortality measures deaths of infants under the age of one, however these deaths are often only reported if the baby is born in a hospital. In many cases however, developing countries have very few hospitals, and thus a significant percentage of babies are born in a more private location, such as in a home. Most likely should a baby that was not born in a hospital die, its death would not be reported, and would contribute to the inaccuracy of that countries infant mortality rate. It is difficult to remove this limitation, but perhaps if countries attempted to collect more accurate health statistics, then better conclusions could be drawn.

A third limitation to this project involved difficulties with technology and transferring files from a MAC computer at home to a PC computer at school. Though the MAC computer does have Microsoft Word and Excel, I still ran into a number of problems that hindered the efficiency of completing my project. The first problem involved excel, as files created on the MAC could not be read properly when opened from a PC. Eventually however, I discovered an alternate way to open the files from a PC at school, which made it possible to create spreadsheets at home and work at them at school. The problem I encountered with Microsoft Word was that on my MAC edition of Word I do not have an equation editor necessary to write complicated formulas such as correlation coefficient and coefficient of determination. Furthermore, at school (though Word *does* have an equation editor) most of the printers can not translate the equations and thus they do not print them properly. This limitation could be removed by using only one type oc computer (preferably PC, as it has an equation editor).

Extensions of Analysis:

This study can act as a basis for further comparisons of the lifestyle and health differences between developing and developed countries. Infant mortality is only one of the possible health indicators of a country, and therefore an extension of this analysis could be to compare number of doctors and/or health care spending to other health indicators such as maternal mortality, life expectancy, child mortality, birth rate, and death rate. Also, if all of the previous factors were to be incorporated, one could produce a very detailed study on the effect of the amount spent on health care and number of doctors in a country on various health indicators.

Oppositely, other factors that may have an effect on a country's infant mortality rate, can be investigated, and hopefully strong correlations can be formed to indicate possible causes of high infant mortality. For instance, the number of midwives could be a variable, as it is the job of midwives to safely deliver babies, whether in a hospital or in a person's home. Infant mortality could also be correlated with the amount of money particular countries spend on health research, thus working towards the prevention of infant death through medication or treatment options.

Both of the previous extensions can be applied separately to developing and developed countries in order to determine not only if correlations can be formed, but also to demonstrate the gap between health care systems and health indicators in developing and developed countries. The results of these studies may illustrate the vast poverty that many countries are facing, and may encourage people in more wealthy, developed nations to donate either money or their services to developing countries in need.

Another extension of this analysis would be to attempt to determine countries that are not strongly developed or strongly developing, but more in between the two. Since there is such a large gap between the wealth and lifestyle of the developing and developed worlds, it would be interesting to test factors against infant mortality in countries in between this gap.

Specifically to the United States, the variation of the infant mortality within states could be validated by also analysing the relative wealth of each state. A relationship

could be (or attempted to be) conjured between the wealth and infant mortality in each state in order to justify each states infant mortality.

Endnotes:

[1] "Least Developed Countries (LDC's)". *United Nations Statistics Division.* 2003. <http://unstats.un.org /unsd/cdb/cdb_dict_xrxx.asp?def_code=481> [09 May 2003].

[2] "World's Second Highest Infant Mortaliy Rate". *Khilafah.com.* (26 October 2002). <http://www.khilafah. com/home/category.php?DocumentID=5425&TagID=2> [18 May 2002].

[3] "Poor Condition of Health Infrastructures Hampers Sectors Growth". *Angola Press Agency.* (30 December 2002). <http://allafrica.com/stories/200212300339.html> [18 May 2003].

[4] "IRC Finds Staggering Infant Mortality Rate in Sierra Leone". *ReliefWeb.com.* (21 February 2001). <http://www.reliefweb.int/w/rwb.nsf/0/f7076e23d6b3fc1 256a000039bb75?Open Document> [17 May 2003].

[5] "Sierra Leone". Kambia Hospital Appeal. <http://www.kambiahospital.org.uk/ information_about/sierra_leone.htm> [17 May 2003].

[6] "Study Shows Primary Health Care More Effective in Villages than Previously Thought". Around the School. (5 May 2000). <http://www.hsph.harvard.edu/ats/May5/> [18 May 2003].

[7] "Tackling Maternal Mortality". Unicef.org. <http://www.unicef.org/noteworthy/ afghanistan/motherhood/index.html [18 May 2003].

[8] ibid

[9] "Maine Infant Morality Remains Very Low". Maine Economic Council. (31 January 2002). <http://www.mdf.org/megc/growth02/measures.php3?m=43> [21 May 2003].

[10] "Mississippi's Infant Mortality Challenge". CNN.com. <http://www.cnn.com/US/9905/ 14/miss.infant.deaths/> [18 May 2003].