**Empirical Results**

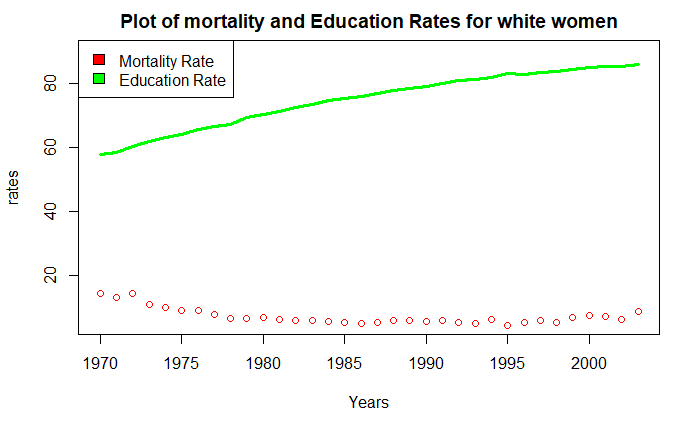
***For white Women***

| **Years** | **Mortality rate** | **Young** | **Old** | **Income** | **Education Rate** |
| --- | --- | --- | --- | --- | --- |
| 1970 | 14.4 | 12989 | 9984 | 45288 | 57.6 |
| 1971 | 13 | 12501 | 9992 | 45941 | 58.4 |
| 1972 | 14.3 | 11970 | 10035 | 47519 | 60.3 |
| 1973 | 10.7 | 11248 | 10130 | 49741 | 61.7 |
| 1974 | 10 | 10980 | 10253 | 52038 | 63 |
| 1975 | 9.1 | 10705 | 10399 | 52936 | 64.1 |
| 1976 | 9 | 13303129 | 9978862 | 55026 | 65.6 |
| 1977 | 7.7 | 13978 | 10131 | 56813 | 66.5 |
| 1978 | 6.4 | 14454 | 10340 | 62695 | 67.2 |
| 1979 | 6.4 | 14791 | 10731 | 69839 | 69.2 |
| 1980 | 6.7 | 15253 | 11049 | 70573 | 70.1 |
| 1981 | 6.3 | 15750 | 11325 | 71566 | 71.2 |
| 1982 | 5.8 | 16323 | 11558 | 71624 | 72.3 |
| 1983 | 5.9 | 16455 | 12282 | 72796 | 73.3 |
| 1984 | 5.4 | 16763 | 12830 | 73977 | 74.6 |
| 1985 | 5.2 | 17063 | 13355 | 74640 | 75.1 |
| 1986 | 4.9 | 17275442 | 13706794 | 75587 | 75.9 |
| 1987 | 5.1 | 17593 | 14308 | 76940 | 76.7 |
| 1988 | 5.9 | 17753 | 14765 | 77493 | 77.6 |
| 1989 | 5.7 | 17836 | 15084 | 77933 | 78.2 |
| 1990 | 5.5 | 17844 | 15524 | 78566 | 79 |
| 1991 | 5.9 | 17746 | 15994 | 78721 | 79.9 |
| 1992 | 5.1 | 17529 | 16543 | 78885 | 80.7 |
| 1993 | 4.9 | 17211 | 16656 | 79484 | 81.3 |
| 1994 | 6.2 | 16862 | 16915 | 80045 | 81.9 |
| 1995 | 4.2 | 16543 | 17165 | 80608 | 83 |
| 1996 | 5.1 | 16339168 | 17354611 | 80741 | 82.8 |
| 1997 | 5.8 | 16041 | 17714 | 81352 | 83.2 |
| 1998 | 5.1 | 15675031 | 17926861 | 82063 | 83.8 |
| 1999 | 6.8 | 15291034 | 18078919 | 83690 | 84.3 |
| 2000 | 7.5 | 14894008 | 18146707 | 84123 | 85 |
| 2001 | 7.2 | 14600083 | 18098667 | 84207 | 85.1 |
| 2002 | 6 | 14341768 | 17939889 | 84014 | 85.2 |
| 2003 | 8.7 | 14226604 | 17658863 | 83852 | 85.7 |

The table above shows mortality rates for white women in America from 1970 to 2003. With various age range which has been classified as Young and Old. The young age group is the combination of 25-29 and 30-34 age groups, while the Old age category is the combination of 35-39 and 40-44 age groups.

The income variable is the number of white women with a means of income in the considered years.

The Education column shows the percentage of white women that are 25 years and over who have completed High school or college by race from 1970 to 2003.



The plot of mortality rate and Education gap widens out as the year progresses. As the education rates increases, so do the mortality rate declines. Only in 1973, 1980, and 1994 did the mortality increased and it did decrease immediately.

**Analysis:**

Performing a multivariate regression analysis model on the data, using mortality rates as the dependent variables and the Young and Old age group with the income variables as the independent variables.

Mathematically, a regression model for the jth sample unit has the form

Yj = β₀ + β₁Zj₁ + β₂Zj₂ + β₃Zj₃ +... βᵣZjᵣ + ᶓ

Where ᶓ is a random error and the βi , i = 0, 1, ..., r are unknown (and fixed) regression coefficients. β0 is the intercept.

On fitting the regression model, we first get a Pearson’s correlation coefficients table for the variables. We have;

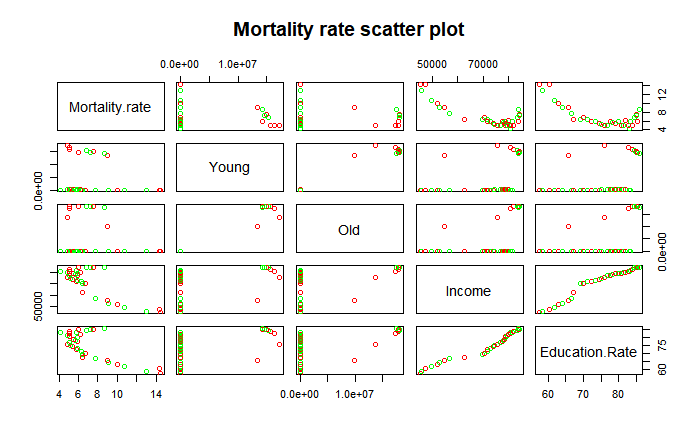
| co-efficients | Mortality Rate | Young | Old | Income | Education Rate |
| --- | --- | --- | --- | --- | --- |
| Mortality Rate | 1.00 | -0.12 | -0.11 | -0.82 | -0.73 |
| Young | -0.12 | 1.00 | 0.98 | 0.40 | 0.48 |
| Old | -0.11 | 0.98 | 1.00 | 0.45 | 0.54 |
| Income | -0.82 | 0.40 | 0.45 | 1.00 | 0.97 |
| Education Rate | -0.73 | 0.48 | 0.54 | 0.97 | 1.00 |

Young age has the strongest relationship with Old age (r = 0.98). Mortality rates are negatively weakly correlated with young age, old age, and negatively strongly correlated with Income and Education rate.

Young age is strongly positive correlated with old ageweakly positively correlated with Income and education rate.

Income is positively correlated with Education rate.

Scatterplots should be produced for each independent with the dependent to see if the relationship is linear. Only Mortality rate, Income and Education rate are the ones showing a linear pattern.



The table above shows the scatter plot for the data.

Call:

lm(formula = Mortality.rate ~ Young + Old + Income + Education.Rate)

Residuals:

| Residuals |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -2.25219 | -0.68607 | -0.00335 | 0.66266 | 2.31955 |

| Co-efficients | Estimate | Std. Error | t value | Pr(>|t|) |
| --- | --- | --- | --- | --- |
| (Intercept) | 1.80E+01 | 3.73E+00 | 4.83E+00 | 4.08e-05 \*\*\* |
| Young | -5.33E-07 | 1.64E-07 | -3.25E+00 | 0.002931 \*\* |
| Old | 5.74E-07 | 1.56E-07 | 3.68E+00 | 0.000957 \*\*\* |
| Income | -3.01E-04 | 7.52E-05 | -4.01E+00 | 0.000393 \*\*\* |
| Number of Educated | 1.36E-01 | 1.18E-01 | 1.159 | 0.255865 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.149 on 29 degrees of freedom

Multiple R-squared: 0.8326, Adjusted R-squared: 0.8095

F-statistic: 36.06 on 4 and 29 DF, p-value: 7.253e-11

Interpretation:

The residuals give a summary of minimum, first quartile, median, third quartile, and maximum values of the model.

The Estimate column in the coefficients table gives us the coefficients for each independent variable in the regression model.

Thus the model becomes:

Mortality rate(y) = 0.18 - 0.000000533(Young) + 0.000000574(Old) - 0.000301(Income)

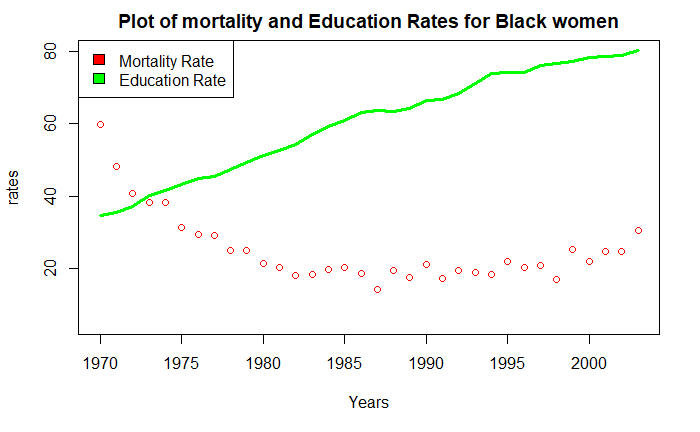
***For Black Women***

| **Years** | **Mortality rate** | **Young** | **Old** | **Income** | **Education Rate** |
| --- | --- | --- | --- | --- | --- |
| 1970 | 59.8 | 1682 | 1330 | 5844 | 34.8 |
| 1971 | 48.3 | 1607 | 1322 | 6151 | 35.4 |
| 1972 | 40.7 | 1540 | 1316 | 6274 | 37.2 |
| 1973 | 38.4 | 1473 | 1313 | 6513 | 40.1 |
| 1974 | 38.3 | 1440 | 1315 | 6779 | 41.5 |
| 1975 | 31.3 | 1402 | 1318 | 6969 | 43.3 |
| 1976 | 29.5 | 1745085 | 1340037 | 7188 | 45 |
| 1977 | 29.2 | 1851 | 1371 | 7562 | 45.4 |
| 1978 | 25 | 1935 | 1404 | 7959 | 47.3 |
| 1979 | 25.1 | 2017 | 1436 | 8533 | 49.2 |
| 1980 | 21.5 | 2116 | 1470 | 8596 | 51.3 |
| 1981 | 20.4 | 2220 | 1507 | 8829 | 52.6 |
| 1982 | 18.2 | 2337 | 1538 | 8921 | 54.3 |
| 1983 | 18.3 | 2430 | 1605 | 9109 | 57.1 |
| 1984 | 19.7 | 2522 | 1673 | 9460 | 59.2 |
| 1985 | 20.4 | 2611 | 1753 | 9611 | 60.8 |
| 1986 | 18.8 | 2677127 | 1822798 | 9819 | 63 |
| 1987 | 14.2 | 2758 | 1936 | 10164 | 63.7 |
| 1988 | 19.5 | 2824 | 2020 | 10380 | 63.4 |
| 1989 | 17.5 | 2861 | 2103 | 10577 | 64.2 |
| 1990 | 21.1 | 2894 | 2201 | 10687 | 66.5 |
| 1991 | 17.2 | 2912 | 2302 | 10727 | 66.7 |
| 1992 | 19.5 | 2913 | 2419 | 11076 | 68.2 |
| 1993 | 19.1 | 2891 | 2506 | 11267 | 71.1 |
| 1994 | 18.5 | 2868 | 2587 | 11450 | 73.8 |
| 1995 | 22.1 | 2837 | 2661 | 11607 | 74.1 |
| 1996 | 20.3 | 2812858 | 2712868 | 11817 | 74.2 |
| 1997 | 20.8 | 2803 | 2790 | 11961 | 76 |
| 1998 | 17.1 | 2787269 | 2855433 | 12272 | 76.7 |
| 1999 | 25.4 | 2762641 | 2915193 | 12383 | 77.2 |
| 2000 | 22 | 2724064 | 2960202 | 12461 | 78.3 |
| 2001 | 24.7 | 2692930 | 2980825 | 12414 | 78.5 |
| 2002 | 24.9 | 2668443 | 2993456 | 12332 | 78.9 |
| 2003 | 30.5 | 2675717 | 2979436 | 12564 | 80.3 |

The table above shows mortality rates for black women in America from 1970 to 2003. With various age range which has been classified as Young and Old. The young age group is the combination of 25-29 and 30-34 age groups, while the Old age category is the combination of 35-39 and 40-44 age groups.

The income variable is the average income for Black women in the considered years in United States dollars. The income data also lags by two years.

The Education column shows the percentage of black women that are 25 years and over who have completed High school or college by race from 1970 to 2003.



A plot of mortality rate and Education is an interesting one. As the education rates increases, so do the mortality rate decreases. It is convenient to say that as people get educated, they adhered and embraced modern medicine. However, for both variables, there were instances of rising and fall.

**Analysis:**

Performing a multivariate regression analysis model on the data, using mortality rates as the dependent variables and the Young and Old age group with the income variables as the independent variables.

Mathematically, a regression model for the jth sample unit has the form

Yj = β₀ + β₁Zj₁ + β₂Zj₂ + β₃Zj₃ +... βᵣZjᵣ + ᶓ

Where ᶓ is a random error and the βi , i = 0, 1, ..., r are unknown (and fixed) regression coefficients. β0 is the intercept.

On fitting the regression model, we first get a Pearson’s correlation coefficients table for the variables. We have;

| co-efficients | Mortality Rate | Young | Old | Income | Education rate |
| --- | --- | --- | --- | --- | --- |
| Mortality Rate | 1.00 | -0.11 | -0.10 | -0.67 | -0.66 |
| Young | -0.11 | 1.00 | 0.99 | 0.55 | 0.56 |
| Old | -0.10 | 0.99 | 1.00 | 0.59 | 1.00 |
| Income | -0.67 | 0.55 | 0.59 | 1.00 | 1.00 |
| Education rate | -0.66 | 0.56 | 0.59 | 1.00 | 1.00 |

Income has the strongest relationship with Education rate (r = 1.00). Mortality rates are negative weakly correlated with young age, old age, Income and Number Educated.

Young age is positive strong correlated with old age, Income and educated rate.

Old is positive moderately correlated with Income and Education rate.

Old age is negative weakly correlated with number educated and negative strongly correlated with Income.



Scatterplots should be produced for each independent with the dependent to see if the relationship is linear. Only mortality rate,Income, and Education rate variables are the ones showing a linear pattern.

Call:

lm(formula = Mortality.rate ~ Young + Old + Income + Education.Rate)

Residuals:

| Residuals |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -8.5214 | -4.493 | -0.1792 | 2.8379 | 18.2687 |

| Co-efficients | Estimate | Std. Error | t value | Pr(>|t|) |
| --- | --- | --- | --- | --- |
| (Intercept) | 7.42E+01 | 7.09E+00 | 1.05E+01 | 2.37e-11 \*\*\* |
| Young | -1.73E-05 | 5.76E-06 | -3.01E+00 | 0.00542 \*\* |
| Old | 2.06E-05 | 5.77E-06 | 3.57E+00 | 0.00127 \*\* |
| Income | -1.30E-02 | 5.98E-03 | -2.17E+00 | 0.03867 \* |
| Education Rate | 1.24E+00 | 8.89E-01 | 1.391 | 0.17475 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.898 on 29 degrees of freedom

Multiple R-squared: 0.6829, Adjusted R-squared: 0.6392

F-statistic: 15.61 on 4 and 29 DF, p-value: 6.381e-07

Interpretation:

The residuals give a summary of minimum, first quartile, median, third quartile, and maximum values of the model.

The Estimate column in the coefficients table gives us the coefficients for each independent variable in the regression model.

Thus the model becomes:

Mortality rate(y) = 0.742 - 0.000073(Young) + 0.000026(Old) - 0.01(Income)

The *R*2 value increases with the number of independent variables. So it is better to use the adjusted R squared value, especially when comparing models. The adjusted R2 indicates that 63% of the variation in Mortality rates can be explained by the model containing Young, Old , and Income.

**Regression Coefficients Comparision**

The research presented in this thesis examines the impact of the female gender mortality rates of white and black races in the United states of America. It relationship with the female population age groups, Income and Education rates.

The previous chapter discussed the use of ordinary (OLS) estimator to determine the slope parameters and to perform hypothesis testing. The two regression models used in this thesis are presented below:

**(V-1)** MRw **=** β₀ + β₁YA + β₂OA + β₃IN + β₄ER + ᶓ

**(V-2)** MRb **=**  β₀ + β₁YA + β₂OA + β₃IN + β₄ER + ᶓ

The slope coefficient, standard error, *R*2, t-statistics, and F- value for each regression equations are presented below with explanations.

**Table V-1**

|  | MRw | t-statistics |
| --- | --- | --- |
| Constant(β₀) | 18.01  (3.73) | 4.83 |
| YA(β₁) | -0.000000533  (0.0000067) | -3.25 |
| OA(β₂) | 0.000000574  (0.000000156) | 3.68 |
| IN( β₃) | - 0.000301  (0.0000752) | -4.01 |
| ER(β₄) | 0.136  (0.118) | 1.159 |
| *R*2 | 0.8095 |  |
| F | 36.06 |  |

The empirical results shown above reveals that if the values of all the slope parameters (β₁...β₄) are equal to zero, the mortality rate for black female in the age groups considered in the united states (MRw) is 18%.

Holding all other independent variables constant, a unit increase in young age population(YA) causes MRw to decrease by 0.000000533%. Whereas,a unit increase in old age population (OA) will cause the the MRw to increase by 0.000000574%. A unit increase in income numbers will decrease MRw by 0.000301% and a unit increase in Education rate will cause MRw to increase by 0.136%.

Going by statistical significance, YA, OA, IN are the only variables that are statistically significant at the 5% significance level. While, ER is statistically insignificant at the 5% significance level.

The *R*2 value increases with the number of independent variables. The adjusted R2 indicates that 63% of the variation in Mortality rates can be explained by the model containing Young, Old , and Income.

The F- statistics generated tests the significance of the regression model.

We can see from our table that Sig. ( p-value) = 6.381e-⁰⁷. From the table,Sig. ( p-value) = 6.381e-⁰⁷, p < 0.05 shows our predictors are significantly better than would be expected by chance. The regression line predicted by the independent variables explains a significant amount of the variance in the dependent variable.

**Table V-2**

|  | MRb | t-statistics |
| --- | --- | --- |
| Constant(β₀) | 74.2  (7.09) | 1.05 |
| YA(β₁) | -0.0000173  (0.0000576 | -3.01 |
| OA(β₂) | 0.0000206  (0.00000577) | 3.57 |
| IN( β₃) | -0.013  (0.00598) | -2.17 |
| ER(β₄) | 1.24  (0.889) | 1.39 |
| *R*2 | 0.6392 |  |
| F | 15.61 |  |

The empirical results shown above reveals that if the values of all the slope parameters (β₁...β₄) are equal to zero, the mortality rate for black female in the age groups considered in the united states (MRw) is 74%.

Holding all other independent variables constant, a unit increase in young age population(YA) causes MRw to decrease by 0.0000173%. Whereas,a unit increase in old age population (OA) will cause the the MRw to increase by 0.0000206%. A unit increase in income numbers will decrease MRw by 0.013% and a unit increase in Education rate will cause MRw to increase by 1.24%.

Going by statistical significance, YA, OA, IN are the only variables that are statistically significant at the 5% significance level. While, ER is statistically insignificant at the 5% significance level.

The *R*2 value increases with the number of independent variables. The adjusted *R*2 indicates that 80% of the variation in Mortality rates can be explained by the model containing Young, Old , and Income.

The F- statistics generated tests the significance of the regression model. From the table, Sig. ( p-value) = 7.253e-¹¹. As p < 0.05 our predictors are significantly better than would be expected by chance. The regression line predicted by the independent variables explains a significant amount of the variance in the dependent variable.