# Regression on Factors Affecting Death Rate

## I. Introduction

Studying mortality rates is one of the unending processes that actuaries are doing as part of their work especially those who are working for life insurance companies. One reason of this continuous study on mortality is the changing world which is inevitable. We play important role in the world's developments, and as a consequence, we may either suffer or enjoy the effects of those changes.

As we define mortality rate, it is the death rate in a population or the number of deaths per thousand people in a population. There are several factors affecting the death rate but in this study, we are going to analyze the impact of hospitalization on the death rate based on a data from city in America.

## II. Methodology

This study is mainly concerned with the effect of three quantitative explanatory variables: Doctor availability, hospital's availability and annual per capita income. The following notations are used for regression analysis:

Response Variable
$Y =$ Death rate

Quantitative Explanatory Variable
$X_1 =$ Doctor availability
$X_2 =$ Hospital availability
$X_3 =$ Annual per capita income

The doctor availability is the number of doctors that can accommodate 1000 people while the hospital availability is the number of hospitals that can accommodate 1000 people. The last factor, per capita income is used in dollar amount.

The data set used in the study is based from the data on US data found in Houghton Mifflin website.

The available data are gathered from cities in America.

Microsoft excel is used to compute for values needed and charts to analyze the quantitative relationships of the variables used.

Multiple regression method is performed to analyze the impact of the explanatory variables to the response variable which is the death rate. The relationship of each factor is modeled as individual variable of linear regression and also as one of variables in multiple regression.

## III. Results and Discussion

Individual relationship on death rate:

For doctor availability, we have the following result using MS excel regression function:

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.178161 |
| R Square | 0.031741 |
| Adjusted R Square | 0.011569 |
| Standard Error | 1.582812 |
| Observations | 50 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 3.942146 | 3.942146 | 1.573527 | 0.215769679 |
| Residual | 48 | 120.2541 | 2.505293 | | |
| Total | 49 | 124.1962 | | | |

| | *Coefficients* | *tandard Err* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 8.407334 | 0.726255 | 11.57628 | 1.7E-15 | 6.947099951 | 9.867568899 | 6.947099951 | 9.867568895 |
| X1 - Doctor Availability | 0.00739 | 0.005891 | 1.254403 | 0.21577 | -0.004454975 | 0.019234403 | -0.004454975 | 0.019234403 |

Here, the data is modeled as follows:

$$Y = 8.407334 + 0.00739X_1$$

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.136252925 |
| R Square | 0.01856486 |
| Adjusted R Square | -0.001881706 |
| Standard Error | 1.593545025 |
| Observations | 50 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 2.305685042 | 2.305685 | 0.90797 | 0.345427711 |
| Residual | 48 | 121.8905158 | 2.539386 | | |
| Total | 49 | 124.1962009 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 8.886285168 | 0.465131189 | 19.1049 | 4.49E-24 | 7.951076249 | 9.821494087 | 7.951076249 | 9.821494087 |
| X2 - Hospital Availability | 0.000642508 | 0.000674284 | 0.952874 | 0.345428 | -0.000713231 | 0.001998246 | -0.000713231 | 0.001998246 |

For the hospital availability, relationship with death rate is modeled here as

$$Y = 8.886285168 + 0.000642508X_2$$

For annual per capita income, we have the following results:

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.132108125 |
| R Square | 0.017452557 |
| Adjusted R Square | -0.003017182 |
| Standard Error | 1.594447786 |
| Observations | 50 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 2.167541239 | 2.167541 | 0.852603 | 0.36043558 |
| Residual | 48 | 122.0286596 | 2.542264 | | |
| Total | 49 | 124.1962009 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 11.15823028 | 2.053033249 | 5.434997 | 1.81E-06 | 7.030330344 | 15.28613021 | 7.030330344 | 15.28613021 |
| X3 - Annual per Capita income | -0.199010378 | 0.215527326 | -0.92336 | 0.360436 | -0.632357103 | 0.234336348 | -0.632357103 | 0.234336348 |

From regression, the data is modeled as:

$$\boldsymbol{Y = 11.15823028 - 0.199010378X_3}$$

Comparing the results above, death rate has almost same relationship with doctor availability and hospital availability based on the regression models. While income effect in the third model is quite different from the first two. However, $R^2$ which measures the goodness of fit is not high for the all the models. Hence, individual effect of each explanatory variable may not be that clear using the linear regression.

For the next analysis, multiple regression model is used in which the three explanatory variables are plot with the death rate all at the same time.

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.294152427 |
| R Square | 0.086525651 |
| Adjusted R Square | 0.026951236 |
| Standard Error | 1.570447484 |
| Observations | 50 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 10.74615708 | 3.582052 | 1.452396166 | 0.239911802 |
| Residual | 46 | 113.4500438 | 2.466305 | | |
| Total | 49 | 124.1962009 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 11.24213426 | 2.076198797 | 5.414768 | 2.15963E-06 | 7.062962906 | 15.42130562 | 7.062962906 | 15.42130562 |
| X1 - Doctor Availability | 0.010694348 | 0.006736979 | 1.58741 | 0.119270664 | -0.002866486 | 0.024255182 | -0.002866486 | 0.024255182 |
| X2 - Hospital Availability | 0.000311536 | 0.00069805 | 0.446294 | 0.657478624 | -0.001093566 | 0.001716638 | -0.001093566 | 0.001716638 |
| X3 - Annual per Capita income | -0.36019862 | 0.234613823 | -1.53528 | 0.131564532 | -0.832451743 | 0.112054507 | -0.832451743 | 0.112054507 |

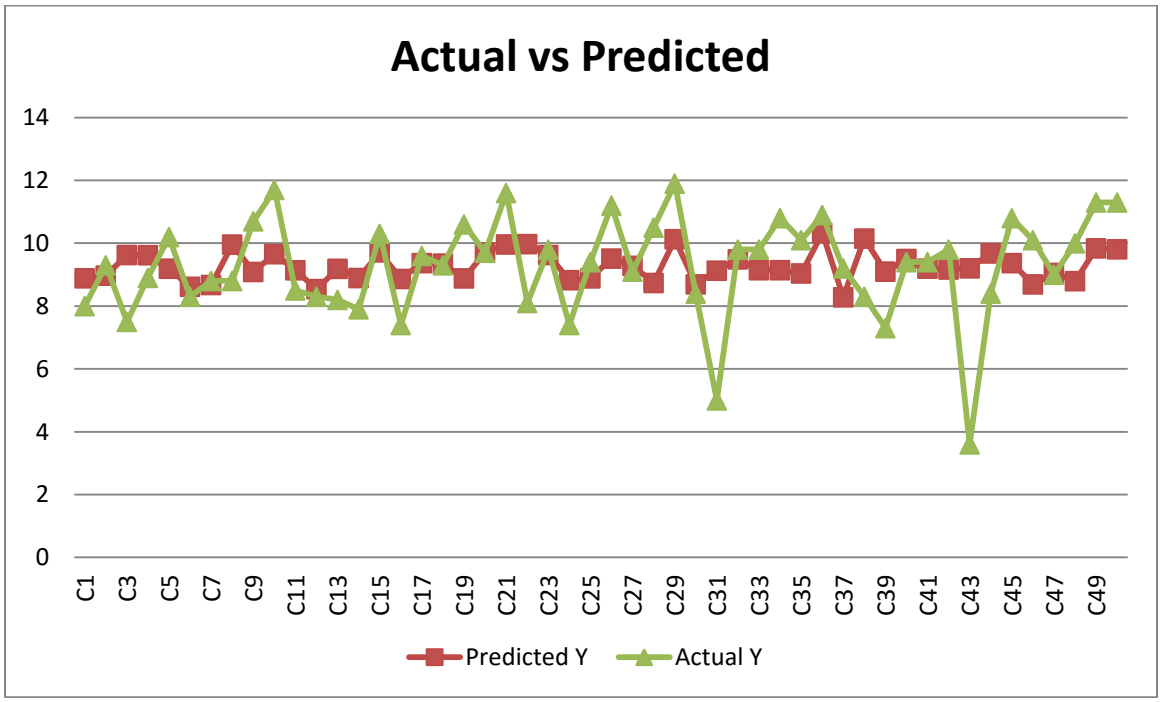
From the result of regression above, the data can be modeled as follows:

$$Y = 11.24213426 + 0.010694348X_1 + 0.000311536X_2 - 0.36019862X_3$$

Based from the resulting model, the $\beta$ coefficients which are basis of quantitative measure of relationship of the explanatory variables with the response variable, $\boldsymbol{X_1}$ can be seen to have highest impact to increase in death rate. It is surprising that doctor availability and hospital availability are measured to have the direct relationship with death rate. On the other hand, we have $\boldsymbol{X_3}$, the annual per capita income that is seen to lower the death rate with its negative coefficient. That is, as the income increases, there are less probable deaths that may occur.

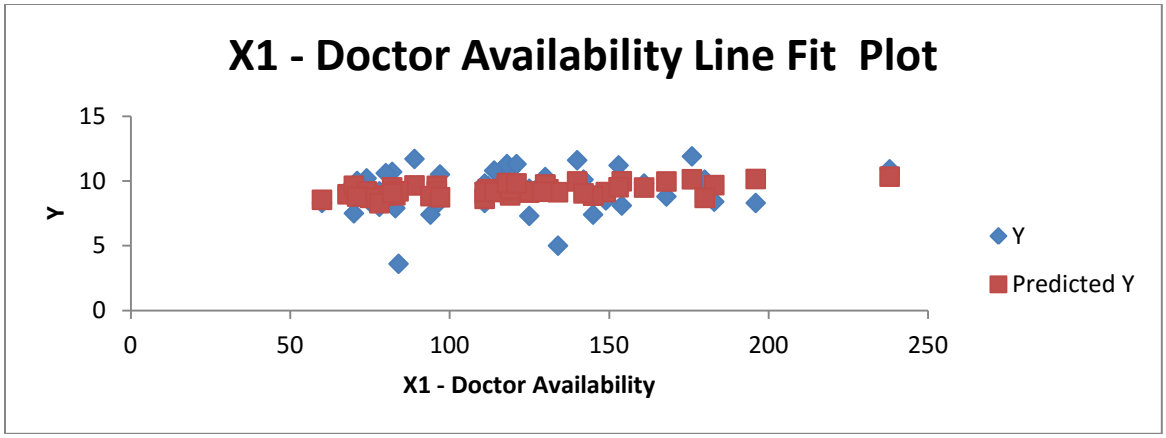
Considering the $R^2$ value, it is not that high to conclude that we have close estimate to the actual.

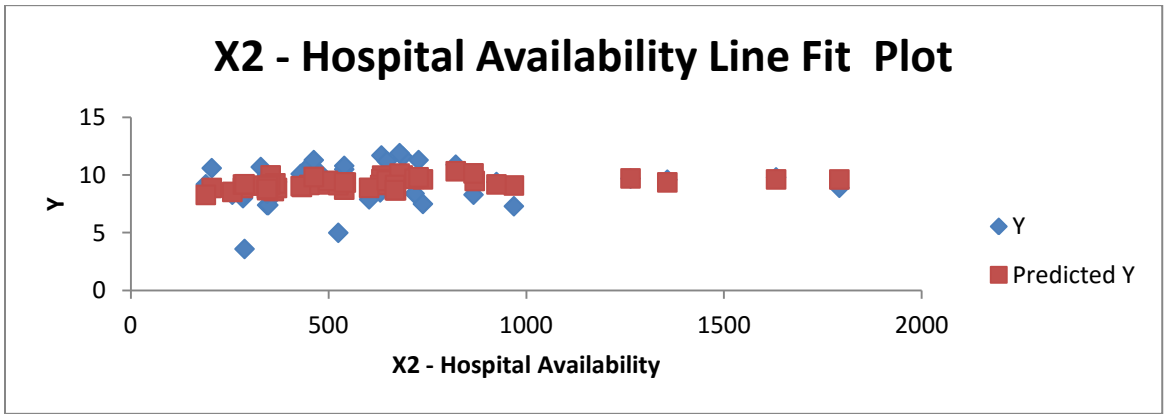
Looking at the predicted values using this model and comparing it with actual, there are fluctuations as can be seen below.

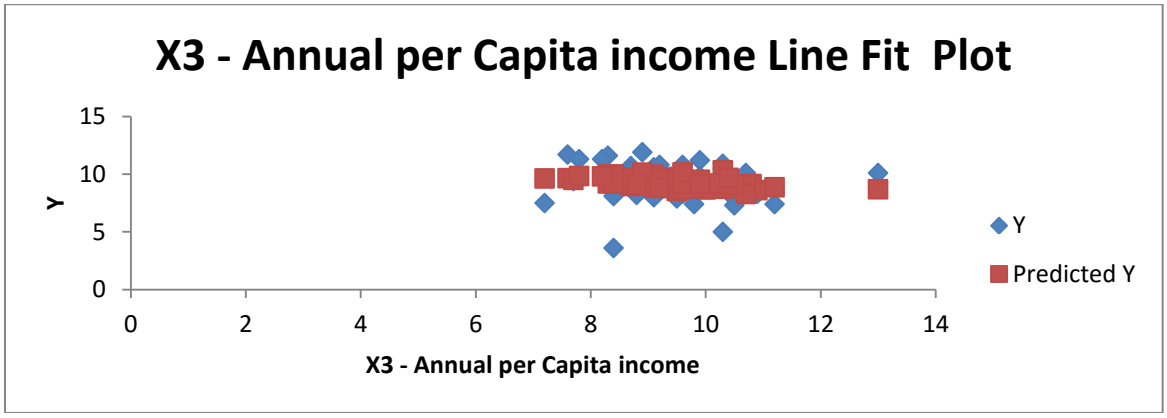



The result of this regression is good only for the data available. The regression analysis performed is largely affected by the limited data gathered. If there will be more data points, the regression fit is expected to be closer to the actual. However for analysis purpose, the model can be acceptable and it already gives ideas on the relationship of the explanatory variables with the response variable.

Plotting the explanatory variables against the actual for, we can see the relationships with the death rate.



## IV. Conclusion

With the concluded model , we have

$$Y = 11.24213426 + 0.010694348X_1 + 0.000311536X_2 - 0.36019862X_3$$

which we can see clearly the different relationship of the three variables with the response variable. Doctor availability and hospital availability is seen to have same moving effect to the death rate but with different level of impacts. For the income per capita, it is seen that it is inversely related with the death rate with its negative sign in coefficient. With this model, fitted regression is just 8.7% close to the actual. This low number may be caused by limitation in available data.

## Tables and References

| Death Rate<br>Y | Doctor Availability<br>X1 | Hospital Availability<br>X2 | Annual Income<br>X3 |
|---|---|---|---|
| 8 | 78 | 284 | 9 |
| 9.3000002 | 68 | 433 | 9 |
| 7.5 | 70 | 739 | 7 |
| 8.8999996 | 96 | 1792 | 9 |
| 10.2 | 74 | 477 | 8 |
| 8.3000002 | 111 | 362 | 11 |
| 8.8000002 | 77 | 671 | 10 |
| 8.8000002 | 168 | 636 | 9 |
| 10.7 | 82 | 329 | 9 |
| 11.7 | 89 | 634 | 8 |
| 8.5 | 149 | 631 | 11 |
| 8.3000002 | 60 | 257 | 10 |
| 8.1999998 | 96 | 284 | 9 |
| 7.9000001 | 83 | 603 | 10 |
| 10.3 | 130 | 686 | 9 |
| 7.4000001 | 145 | 345 | 11 |
| 9.6000004 | 112 | 1357 | 10 |
| 9.3000002 | 131 | 544 | 10 |
| 10.6 | 80 | 205 | 9 |
| 9.6999998 | 130 | 1264 | 9 |
| 11.6 | 140 | 688 | 8 |
| 8.1000004 | 154 | 354 | 8 |
| 9.8000002 | 118 | 1632 | 9 |
| 7.4000001 | 94 | 348 | 10 |
| 9.3999996 | 119 | 370 | 10 |
| 11.2 | 153 | 648 | 10 |
| 9.1000004 | 116 | 366 | 9 |
| 10.5 | 97 | 540 | 10 |
| 11.9 | 176 | 680 | 9 |
| 8.3999996 | 75 | 345 | 10 |
| 5 | 134 | 525 | 10 |
| 9.8000002 | 161 | 870 | 10 |
| 9.8000002 | 111 | 669 | 10 |
| 10.8 | 114 | 452 | 10 |
| 10.1 | 142 | 430 | 11 |
| 10.9 | 238 | 822 | 10 |
| 9.1999998 | 78 | 190 | 11 |
| 8.3000002 | 196 | 867 | 10 |
| 7.3000002 | 125 | 969 | 11 |
| 9.3999996 | 82 | 499 | 8 |
| 9.3999996 | 125 | 925 | 10 |
| 9.8000002 | 129 | 353 | 10 |
| 3.5999999 | 84 | 288 | 8 |
| 8.3999996 | 183 | 718 | 10 |
| 10.8 | 119 | 540 | 9 |
| 10.1 | 180 | 668 | 13 |
| 9 | 82 | 347 | 9 |
| 10 | 71 | 345 | 9 |
| 11.3 | 118 | 463 | 8 |
| 11.3 | 121 | 728 | 8 |

http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/frames/frame.html