**Analysis of Australian AIDS Data:**

**A Logistic Regression Modeling Approach**

**Patryk Siczek  
Sebastian Nasterski**

**Introduction**

Globally, the human immunodeficiency virus (HIV) is a leading cause of morbidity and mortality. In 2018, about 37.9 million people were living with HIV globally, with a disproportionate 70% living in sub-Saharan Africa. Over the years, effective antiretroviral therapies (ARTs) have evolved, thereby increasing the life expectancy and quality of life of HIV-infected individuals. The global community, in its effort to alleviate HIV infections, set an ambitious goal dubbed the “90-90-90 policy” in 2014. This initiative seeks to ensure that by 2020, 90% of people living with HIV are aware of their HIV status, 90% of those diagnosed with HIV have access to ART, and 90% of those receiving ART achieve suppression of viral loads. Such concerted global efforts have shifted the survival with HIV infection from a terminal illness to a chronic condition.

Although the introduction of antiretroviral therapy (ART) has substantially reduced hospitalization among people living with human immunodeficiency virus ([HIV] PLWH), hospitalizations among this group have still been reported to be elevated compared to individuals hospitalized with conditions other than HIV. In fact, HIV-infected individuals have been twice as likely to experience hospitalization, and experience higher in-hospital mortality rates than the general inpatient population. Although over the past several decades, this rate has decreased substantially, partly due to the introduction of ART and the reduction of acquired immune deficiency syndrome (AIDS), hospitalization rates among PLWH remain high, which may be extremely costly and burdensome for hospitals and insurance companies. Recent research has also indicated that hospitalization rates among PLWH have been underestimated, which can erroneously mask the magnitude of suffering among this population.

While research suggests that hospitalizations among PLWH are largely manageable, longer duration of hospital stays are more common among this population compared to their uninfected counterparts. Prolonged hospital stays are also commonly associated with hospital-acquired infections among those already affected by infectious diseases. Longer stays in the hospital place a substantial burden on hospital and health insurance systems, typically requiring more resources for treatment. In this study, the objective is to determine the how length of stay in the hospital impact the survival of the diagnosed AIDS patients alongside some identified demographic information.

**Methods**

The model used for this work is stated as follow

Where the p is the proportion of the survived patient after the diagnosis. The test associated the model is evaluated at 5% level. Also, the preliminary analysis was also carried which serves as the prior analysis and uncover other knowledge about the study.

**Data and Analysis**

To achieve the aim of the study, Australian AIDS data were collected from <https://vincentarelbundock.github.io/Rdatasets/datasets.html> with data descriptions available at <https://vincentarelbundock.github.io/Rdatasets/doc/MASS/Aids2.html>. The data contains a total of 2843 patients diagnosed with AIDS. 1084 patients were found to survive the disease after the diagnosis which accounts for a total of while 1761 resulted to death, out of these diagnosed patients, 2754 were male and total female cases were found to be 89. It was obtained that 1780 cases were from NSW while 1063 are from other areas in Australian which suggest that about two-third of the cases of AIDS found in Australia are from NSW. The average of an individual with AIDS were found to be 37.41 (SD=). The length of stay in the hospital after the diagnosis was computed, the result shows that overall length of states in the hospital were computed to be 405.93 (SD=) which is about 14 months after the diagnosis was carried out. The preliminary analysis shows that the age and length of stay in the hospital by the diagnosed patients have outliers, as such the data was subjected to preprocessing in a bid to remove the outliers present in the data set. The result of the processed data is presented below;

Table 1: Descriptive Analysis

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Overall (n=2693) | | Survive (n=1021) | | Dead (n=1672) | |
| Mean | SD | Mean | SD | Mean | SD |
| Age | | 36.99 | 8.67 | 36.64 | 8.62 | 37.21 | 8.70 |
| LOS | | 379.86 | 304.32 | 430.56 | 338.24 | 348.89 | 277.18 |
|  |  | Freq | Percent | Freq | Percent | Freq | Percent |
| Location | NSW | 1680 | 62.38 | 622 | 60.92 | 1058 | 63.28 |
|  | Others | 1013 | 37.62 | 399 | 39.08 | 614 | 36.72 |
| Sex | Male | 2624 | 97.44 | 992 | 97.16 | 1632 | 97.61 |
|  | Female | 69 | 2.56 | 29 | 2.84 | 40 | 2.39 |

Australian AIDS Survival Data, Source <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

The result in table 1 shows that the realized usable cases for this work after the preprocessing was found to be 2693 with a total of 1021 (37.9%) survived cases and a total of 1672 (62.1%) death cases due to AIDS. It was obtained that average age (Mean=36.64, SD=8.62) of the survived AIDS patients was relatively lower than that of cases resulted to death (Mean=37.21, SD=8.70) with the mean length stay of the survived patients (Mean=430.56, SD=338.24) higher than that of the death cases (Mean=348.89, SD=277.18). 622 (60.92%) of the survived AIDS patients are from NSW while 399 (39.08%) of the survival are located in other part of Australia and the cases resulted to have 1058 (63.28%) located in the NSW part of Australia while 614 (36.72%) came from other part of the country. This suggest that survival are less frequent in NSW as compared to the AIDS related death and likewise, survival are more frequent in other part of the country than the recorded death. It is obtained that male gender dominates the population of the AIDS cases with more male patient resulting to death than the female patients.

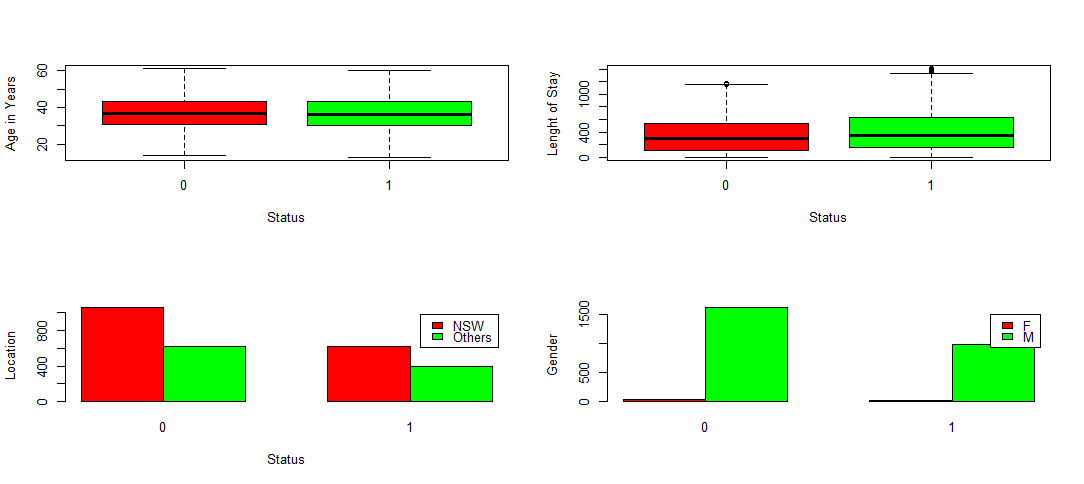


Figure 1: Pictorial Description to the data presented in table 1

We then proceed to test hypothesis using the univariate approach. The result of the test in table 2 featured the chi-squared test of independent and the t-test. The test proceed to determine if the survival of AIDS patients are dependent of location and as well, if the survival are dependent on the sex of the patients under study. The result of the test in table 2 (lower part) determine if the average age and length of stay in the hospital are the same between the cases of AIDS patients status. The result of the test showed that the AIDS survival are statistically insignificant of the location of the patients (chisq=1.402, df=1, p=0.236), the same was also found for sex (chisq=0.346, df=1, p=0.556). The result of the test of difference based on the t-test was found to be statistically insignificant for age (t=1.654, df=2170.8, p=0.098) while the length of stay in the hospital are significantly different (t=-6.497, df=1839.3, p= <0.001) between the AIDS patients who survive or those result to death after the diagnosis.

Table 2: Simple Analysis and Test of Hypothesis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Chi-Squared Test | Status | | Chi-Squared | df | p-value |
|  | Death | Survived |
| Location | NSW | 1058 (1043.06) | 622 (636.94) | 1.402 | 1 | 0.236 |
| Others | 614 (628.94) | 399 (384.06) |  |  |  |
| Sex | Female | 40 (42.84) | 29 (26.16) | 0.346 | 1 | 0.556 |
|  | Male | 1632 (1629.16) | 992 (994.84) |  |  |  |
|  | T-Test | Mean | | t-value | df | p-value |
|  | Age | 37.210 | 36.642 | 1.654 | 2170.8 | 0.098 |
|  | LOS | 348.894 | 430.558 | -6.497 | 1839.3 | <0.001 |

The result of the logistic regression model is presented in table 3 alongside the deviance parameter and model accuracy. The result of the fitted model shows that only the length of stay in the hospital; after the diagnosis was found to be statistically significant at 5% level. The odds ratio was also computed by taking the exponent of the estimated coefficient. The result of the odds ratio shows that for every unit increase in the age and length of stay in the hospital, the odds of survival after the diagnosis will increase by 0.99577 times and 1.00086 times respectively. The odds associated with the dummy variable location was found to be 1.10864, this tells that patients from NSW are 1.10864 less likely to survive after the diagnosis. Likewise, the male patients are 0.8808 times less likely to survive than the female patients. The deviance parameter associated with the fitted model was computed to be 1.3119 which shows that the dispersion parameter on the fitted model was relatively. The model accuracy was computed as 0.6405 which means that the model correctly predict 64.05% of the outcome variable. The ROC curve was presented alongside the AUC.

Table 3: Logistic Regression Model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Coefficients | Std Error | z-value | p-value | Odds |
| Intercept | -0.5856 | 0.307 | -1.909 | 0.0562 | 0.55668 |
| Age | -0.004 | 0.005 | -0.908 | 0.364 | 0.99577 |
| LOS | 0.001 | 0.0001 | 6.542 | <0.0001 | 1.00086 |
| Location (Others) | 0.103 | 0.0825 | 1.25 | 0.2114 | 1.10864 |
| Sex (Male) | -0.127 | 0.2508 | -0.506 | 0.6128 | 0.8808 |
| Dispersion Parameter | 1.3119 |  |  |  |  |
| Model Accuracy | 0.6405 |  |  |  |  |

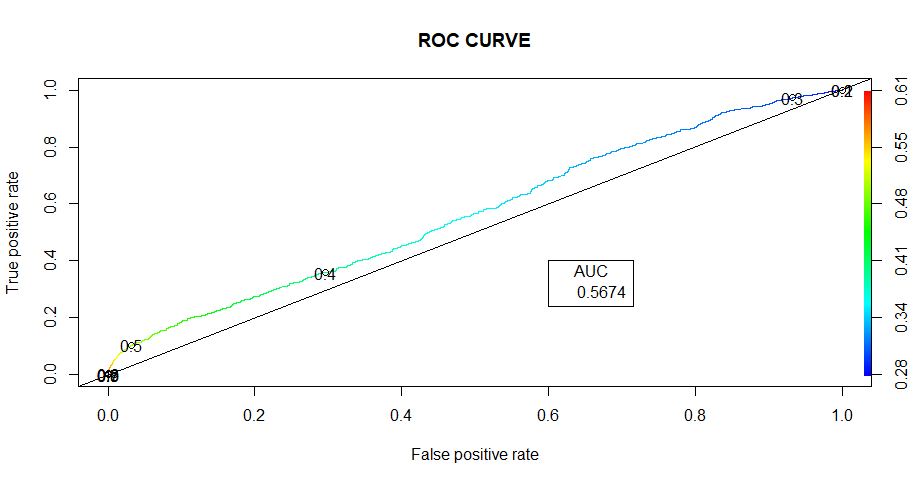


Figure 2: AUC-ROC plot

**Conclusion**

The result finds that length of stay in the hospital have a significant effect on the survival AIDS patients while the identified demographic ware found to be insignificant. The result shows that for one day additional stay in the hospital after the diagnosis, the odds of survival will increase by 0.086%. Also, the older the patients the less the survival of the AIDS patients after the diagnosis. Also, Male patients are less likely to survive than the female patients.