# Abstract

Machine learning has been widely used in analyzing and interpreting electronic healthcare records. The integration of machine learning could improve healthcare decisional making and support the diagnostic operation. Yet predictive models may also induce unexpected consequences. One of them is indirect discrimination. The machine learning algorithm is known as objectively discover pattern and make prediction. However, data-driven predictive models may end up with discriminate certain groups of people and worsen social disparity. In this study, we review a predictive model for heart transplantation survival probabilities and introduce statistic disciplines to measure discrimination. This survey is primarily intended for researchers to utilize data mining and machine learning to develop a non-discrimination predictive model. In addition, practitioners and policymakers would use the study for diagnosing potential discrimination by predictive models.

Keywords: bias, decision making, AI, machine learning, heart transplantation

# Introduction

Increasingly, machine learning has been widely adopted in critical applications where decisions are derived from complex data. (Mozaffari-Kermani, Sur-Kolay, Raghunathan, & Jha, 2015) In particular, the healthcare decision-making process involves models established in interpreting and analyzing complex historical data. (Jenna Wiens, 2018) Meanwhile, as data suggested, the use of electronic healthcare data has increased dramatically in the last five years. (Benjamin A Goldstein, January 2017) With the increasing pervasiveness of electronic health data and the need to process and interpret complex historical data, machine learning has played a significant role in healthcare data handling and analysis. (Shenoy, 21 August 2017)

Machine learning built on healthcare data can generate actionable insights from improving patient care to predicting potential disease. Meanwhile, the data-driven predictive model could assist physicians and staff members who provide healthcare targeted information with diagnostic support or advanced analytics information that improves clinical decision making or offer diverse treatment options. (Milena A. Gianfrancesco, November 2018)

As machine learning becomes taking more roles in decision support, many have pointed out the potential risk. Some risks around fairness and discrimination are that even if the algorithm is fair and well-intentioned, the integration of machine learning may discriminate against certain groups of people. (Žliobaitė, 2017) Such bias or discrimination could be unintentional or unexpected yet severe. (Žliobaitė, 2017) Furthermore, the penetration of machine learning in other fields outside healthcare has already aroused people’s concern for social or economic disparities such as legal and justice systems, advertisements, or computer vision. (Milena A. Gianfrancesco, November 2018)

## Definition of Discrimination

Discrimination refers to an unjustified treatment of patients on the basis of gender or race. (Ana I Balsa, 2005) Human rights laws prohibit discrimination on the grounds of race, national or ethnic origin, color, religion, age, sex, sexual orientation, gender identity or expression, marital status, family status, genetic characteristics, or disability. (RUGGIERI, 2013) Research studies for disparity have diverged focus on different areas such as health care, mortgage, or education. In particular, as an advanced tool for medical decision support, the machine learning algorithm has received lots of publicity along with the operation of electronic healthcare records. A promise of machine learning in health care is the circumvent of biases in treatment. Yet, the objectivity of the algorithm remains questioned, and the examination of the model is required.

## Discrimination in Machine Learning

Discrimination is usually divided into two categories in the legal file: direct discrimination and indirect discrimination. Direct discrimination refers to the situation where an individual receives less favorable treatment based on their protected attributes. For instance, a qualified black man could be rejected for his mortgage application just because of his race. Indirect discrimination occurs when individuals are treated based on non-protected attributes but still unjustified by their protected attributes. A famous example would be redline, where service was limited for residents of specific areas based on race. Although location is neutral attribute, it correlates with ethnicity as demographic makeup.

Machine learning bias and discrimination could generate systematic errors that induce unfair consequences. In particular areas such as banking, job market, and medicine, the process may be heavily tilted as consequential decisions are often informed by statistical risk assessments that quantify the possibility of potential courses of action. (Sam Corbett-Davies, 2018) Examples are not hard to seek; in 2015, the hiring algorithm that Amazon designed to review the resume and assist the human resource department was canned a year later once they discovered tools systematically discriminated against women applying for technical jobs, such as software engineer positions. In October 2019, Optum sold an algorithm to guide care decision-making for millions of people. The algorithm was identified as biased later as the system heavily privileged white people over black people. (Jee, 2019) It is clear that the predictive model's bias has become one of the social issues that need to be addressed immediately.

## Protected Groups of Gender and Region

We choose to use the heart transplantation predictive model as heart failure is a worldwide pandemic, and heart transplantation is the most effective treatment for patients with end-stage heart failure. The model is designed to predict heart transplantation survival status and survival possibility based on patients' information. 0 in survival status means the patient would not survive, and 1 in survival status indicates the patient would survive. Meanwhile, the survival possibility is ranged from 0 to 1.

The algorithm evaluates patients' information and estimates whether the patient would survive. However, the features that platform analyzed also includes some protected class such as gender and region. The gender is displayed as male and female while the region is split into Southeast, Middle west, and Northeast. During the computation, the algorithm may favor certain groups of people in those two protected groups. For example, males could end up with higher survival possibilities, or people from certain regions may be discriminated against with lower survival possibilities. We used the following methodology to test our supposition.

# Methodology

To investigate potential bias and discrimination in a typical AL platform for heart transplantation, we propose a statistical tests approach. There are four common ways for measuring discrimination. The first approach involves the use of statistical tests to check how likely the observed result is to have occurred by chance. The second approach is absolute measures. Absolute measures aim to detect the magnitude of differences between groups of people. Conditional tests are used in the third approach to identify how much of difference between the groups. They are designed for indicating the presence/ absence or magnitude of indirect discrimination. On the other hand, for quantifying direct discrimination, structure measures can be used by checking and analyze each individual is discriminated and how many of individual affected.

This paper focus on two main parts that consists of a. regression slope test b. difference of means test. The details of each statistical tests’ methods are organized as follows.

## Data Source

The dataset used in this study has been provided by UNOS which covered 103,570 heart transplant events. These datasets processing was created by Professor Hamid team, include: (a) features from year 1 to year 10 (b)actual survival results (c)predict survival results.

## Regression slope test

The Regression slop test approach describes how to perform hypothesis test to determine whether there is a significant linear relationship between the independent variable X and the dependent variable Y. The test conforms to an ordinary least squares (OLS) regression to the protected variable. In order to investigate discrimination, it is effective to test whether the regression coefficient of the protected variable is significantly different from zero. This test focus on the slope of the regression line:

y (actual or predicted value) = b\*s (protected variable) + c (constraint) (where b is the estimated regression coefficient of protected variable)

The first procedure of regression slope test is to state the hypotheses:

H0: b=0

Ha: b≠ 0

If there is a significant linear relationship between the protected variable and actual or predictive value, the slope will not equal to 0. Otherwise, the slope is equal to 0.

In general, the protected group and the target variable can be either binary or numeric. This study just investigates the more common scenarios: the protected group in the discrimination testing is binary. A linear regression t-test: t=b/σ can be used as the test statistics formulation, where σ is the standard error and can be computed as σ = , b indicates the estimated regression coefficient of protected group variables, f(.) indicates the regression model, ̄. is the mean. the degrees of freedom (DF) are equal to DF = n – 2, where n is the number of observations in the sample.

## Difference of Means Test

This phase provides a short overview of the difference of means test. It is much more common to conduct a hypothesis test for the difference of means than the specific values of the means themselves. This section covers how to test for the difference between two means from two separate groups and how to conduct the analysis of variance (ANOVA) to compare means of a certain variable in two and more independent groups.

*Hypothesis test for the difference between two means*

Initially, the test assumes the two groups have the equal variances, the groups are normally distributed, and each value is sampled independently. The null and alternative hypothesis are often stated as following:

Null hypothesis: Ho: μ1 = μ2(there is no difference between the two groups means)

Alternative hypothesis: Ha: μ1 ≠ μ2

The test statistic formulation is t=, where and indicate the number of individuals in the regular group and the protected group respectively , σ computed equal to , where is the sample variance in the regular group and are the target variance in the protected group. DF= is applied to the t-test.

*One Way Analysis of variance (ANOVA) for difference between two and more means*

The ANOVA technique applies to test for a difference means in more than two independent groups. The procedure of ANOVA is conducted using the same several steps we discussed in previous section. The null hypothesis states that there is no difference between the three groups means. The null and alternative hypothesis are often stated as following:

Null hypothesis: Ho: μ1 = μ2 = … = μk

Alternative hypothesis: The means are not all equal.

The test statistic: , where the is the sample size in the group, k is the number of independent groups, indicates the overall mean, and represents the sample mean in the group.

In One Way ANOVA test, hypothesis captures any difference in means. For example, we specify three groups: Southeast, Middle west, and Northeast​ as regions, and the region where all three means are unequal, where one is different from the other two, where two are different, and so on. The alternative hypothesis, as shown above, capture all possible situations other than equality of all means specified in the null hypothesis.

In this paper, the described methodologies mentioned as previous can be deployed to gender group and region group respectively. First, the regression slop test and mean differences test are applied to gender groups see if there is significant difference among protected group and regular group for actual survival status as well as for the predicted survival status and predicted survival possibilities. Then the regression slope tests are performed to region group by repeating the same procedure. However, One Way ANOVA applied to region group for mean difference test. As for the results, the reader should note that the null hypothesis should be accepted if the P-values are greater than 0.05.

# Result and Discussion

For each of ten years the study covered, the data was divided into protected groups according to gender and region. Regarding the protected groups, the statistical tests results of actual survival rate and predicted survival rate, and the statistical test results of actual survival rate and the predicted survival possibilities, were compared in terms of critical values and p-values. Both regression slop test and difference of means test are performed.

## Test Results in Terms of Gender

### Result of Regression Slop Test

The gender was used to perform the regression slop test, where the gender as protected feature and the actual survival rate as the target. The result shows that all p-values are greater than 0.05 for all the ten years in actual survival status as concluded from Table 1, which indicates that the null hypothesis has to be accepted. There is no significant linear relationship between gender and actual survival status.

However, as shown in Table 2, the p-values generated by regression slop tests where gender as protected feature and the predicted survival rate as the target are less than 0.05 for all ten years. The null hypothesis is therefore rejected. The significant linear relationship exists between gender and predicted survival status.

The existence of significant relationship between gender and predicted survival probabilities is also proved as shown in Table 3.

**Table 1.** Results including Slope, Standard Error, P-value, and conclusion of regression slop test performed for gender and actual survival status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Actual Survival Status | | | | |
| Year | Slope | Standard error | P-value | Conclusion |
| 0 | 0.001510795 | 0.006604671 | 0.819070857 | Accept |
| 1 | 0.012958355 | 0.009701489 | 0.181678528 | Accept |
| 2 | 0.002991531 | 0.011006689 | 0.785788459 | Accept |
| 3 | 0.013493148 | 0.011837513 | 0.25437692 | Accept |
| 4 | 0.013493148 | 0.011837513 | 0.25437692 | Accept |
| 5 | 0.012716682 | 0.013391091 | 0.342327845 | Accept |
| 6 | 0.022472543 | 0.014056755 | 0.109932553 | Accept |
| 7 | 0.009209608 | 0.014481582 | 0.524828878 | Accept |
| 8 | 0.001508151 | 0.014928041 | 0.919531364 | Accept |
| 9 | 0.006533696 | 0.015152775 | 0.666346925 | Accept |
| 10 | 0.018310947 | 0.015338018 | 0.23259407 | Accept |

**Table 2.** Results including Slope, Standard Error, P-value, and conclusion of regression slop test performed for gender and predicted survival status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Status | | | | |
| Year | Slope | Standard error | P-value | Conclusion |
| 0 | 0.050847491 | 0.011930913 | 2.05E-05 | Reject |
| 1 | 0.035386855 | 0.012385337 | 0.004284322 | Reject |
| 2 | 0.038522079 | 0.01273741 | 0.002499292 | Reject |
| 3 | 0.045311013 | 0.013024613 | 0.00050617 | Reject |
| 4 | 0.045311013 | 0.013024613 | 0.00050617 | Reject |
| 5 | 0.089718418 | 0.013694132 | 6.08E-11 | Reject |
| 6 | 0.077485842 | 0.014078576 | 3.85E-08 | Reject |
| 7 | 0.052495822 | 0.014398259 | 0.000268411 | Reject |
| 8 | 0.047807877 | 0.014736314 | 0.001183829 | Reject |
| 9 | 0.034138781 | 0.014957594 | 0.022501932 | Reject |
| 10 | 0.04243257 | 0.015247362 | 0.005404143 | Reject |

**Table 3.** Results including Slope, Standard Error, P-value, and conclusion of regression slop test performed for gender and predicted survival probabilities

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Probabilities | | | | |
| Year | Slope | Standard error | P-value | Conclusion |
| 0 | 0.019286084 | 0.004130946 | 3.07E-06 | Reject |
| 1 | 0.014649565 | 0.003554636 | 3.80E-05 | Reject |
| 2 | 0.011839077 | 0.003172059 | 0.000190972 | Reject |
| 3 | 0.014741208 | 0.003082325 | 1.76E-06 | Reject |
| 4 | 0.020766966 | 0.003106283 | 2.46E-11 | Reject |
| 5 | 0.028700708 | 0.003402421 | 3.94E-17 | Reject |
| 6 | 0.021556139 | 3.41E-10 | 0.003428245 | Reject |
| 7 | 0.018114143 | 0.003742111 | 1.32E-06 | Reject |
| 8 | 0.017326496 | 0.0045023 | 0.000120078 | Reject |
| 9 | 0.01385538 | 0.005581601 | 0.013079482 | Reject |
| 10 | 0.015449898 | 0.00665848 | 0.020357182 | Reject |

### Result of Difference of Means Test

The data was divided into male group and female group, and the difference of means tests were performed between the two protected groups for each of actual survival status, predicted survival status, and predicted survival probabilities.

Table 4 concludes the results of difference of means tests comparing the mean of actual survival status of male group and the mean of actual survival status of male group. For all ten years, The P-values are greater than 0.05, which suggests that the null hypothesis should be accepted. The means of actual survival rates of male group are not significant different from that of male group.

The opposite results are concluded in Table 5 for predicted survival status. For most years (except for year 1), the hypothesis is rejected since the p-values are much less than 0.05. Even for year 1, the p-value, approximately 0.07, is close to 0.05. In the most cases, means of predicted survival rates of male group are significant different from that of male group.

Similarly, for predicted survival probability, the difference of means tests provide the conclusion that the significant difference exists between male and female group regarding predicted survival probability since all the p-values are less than 0.05 as shown in Table 6.

**Table 4.** Results including statistic, p-value, and conclusion of mean difference test performed for actual survival status of male group and female group.

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Survival Status | | | |
| Year | Statistic | P-value | Conclusion |
| 0 | 0.2287465 | 0.819070857 | Accept |
| 1 | 1.119031273 | 0.263172924 | Accept |
| 2 | 0.271792097 | 0.785788459 | Accept |
| 3 | 1.139863429 | 0.25437692 | Accept |
| 4 | 1.575004988 | 0.115296461 | Accept |
| 5 | 0.949637501 | 0.342327845 | Accept |
| 6 | 1.598700698 | 0.109932553 | Accept |
| 7 | 0.635953189 | 0.524828878 | Accept |
| 8 | 0.101028084 | 0.919531364 | Accept |
| 9 | 0.431188121 | 0.666346925 | Accept |
| 10 | 1.193827451 | 0.23259407 | Accept |

**Table 5.** Results including statistic, p-value, and conclusion of mean difference test performed for predicted survival status of male group and female group.

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted Survival Status | | | |
| Year | Statistic | P-value | Conclusion |
| 0 | 4.261827396 | 2.05E-05 | Reject |
| 1 | 1.760658087 | 0.078348741 | Accept |
| 2 | 3.024325933 | 0.002499292 | Reject |
| 3 | 3.478876085 | 0.00050617 | Reject |
| 4 | 5.693437819 | 1.29E-08 | Reject |
| 5 | 6.55159577 | 6.08E-11 | Reject |
| 6 | 5.503812349 | 3.85E-08 | Reject |
| 7 | 3.645983962 | 0.000268411 | Reject |
| 8 | 3.244222288 | 0.001183829 | Reject |
| 9 | 2.282371068 | 0.022501932 | Reject |
| 10 | 2.78294493 | 0.005404143 | Reject |

**Table 6.** Results including statistic, p-value, and conclusion of mean difference test performed for predicted survival probabilities of male group and female group.

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted Survival Probabilities | | | |
| Year | Statistic | P-value | Conclusion |
| 0 | 4.668684187 | 3.07E-06 | Reject |
| 1 | 2.729503869 | 0.00636202 | Reject |
| 2 | 3.732300803 | 0.000190972 | Reject |
| 3 | 4.782496659 | 1.76E-06 | Reject |
| 4 | 6.685470381 | 2.46E-11 | Reject |
| 5 | 8.435379497 | 3.94E-17 | Reject |
| 6 | 6.287806606 | 3.41E-10 | Reject |
| 7 | 4.84062195 | 1.32E-06 | Reject |
| 8 | 3.848365066 | 0.000120078 | Reject |
| 9 | 2.482330696 | 0.013079482 | Reject |
| 10 | 2.320333933 | 0.020357182 | Reject |

### 

## Test Results in Terms of Region

### Result of Regression Slop Test

As introduced in the 2.4., the data were separated according to three regions – middle west, north west, and north east – to perform the regression slop test which examine the linear relationship between region and targets, including actual survival status, predicted survival status, and predicted survival probabilities. The details of results for middle west, north east, and south east region are displayed correspondingly in Table 7, Table 8, and Table 9.

In fact, all the three tables provide a similar indication, that the hypothesis is mostly accepted when actual survival status is tested as target, yet the hypothesis is mostly rejected when predicted survival status and predicted survival probabilities are tested as target. Therefore, there is no significant linear relationship between region and actual survival status; however, there do exist significant linear relationship between region and predicted values, including predicted survival status and predicted survival probability.

**Table 7.** Results including Slope, Standard Error, P-value, and Conclusion of regression slop test performed for Midwest region and the set of actual survival status, predicted survival Status, and predicted survival probabilities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Actual Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -0.00575387 | 0.376377189 | 0.006504249 | Accept |
| 1 | 0.011935139 | 0.208334765 | 0.009485524 | Accept |
| 2 | 0.016821817 | 0.117115214 | 0.010733974 | Accept |
| 3 | 0.015116247 | 0.196110337 | 0.011692464 | Accept |
| 4 | 0.023837115 | 0.055766259 | 0.012459631 | Accept |
| 5 | 0.021161617 | 0.105788922 | 0.013081994 | Accept |
| 6 | 0.021985872 | 0.108196539 | 0.013685069 | Accept |
| 7 | 0.023311525 | 0.100970746 | 0.014210933 | Accept |
| 8 | 0.03227824 | 0.026408173 | 0.014535323 | Reject |
| 9 | 0.021004991 | 0.15625686 | 0.014813636 | Accept |
| 10 | 0.019313073 | 0.198194901 | 0.015007928 | Accept |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | 0.058055871 | 7.84E-07 | 0.011746106 | Reject |
| 1 | 0.054161665 | 7.71E-06 | 0.012101536 | Reject |
| 2 | 0.072222822 | 6.03E-09 | 0.012405586 | Reject |
| 3 | 0.088541746 | 5.70E-12 | 0.012837269 | Reject |
| 4 | 0.122214928 | 1.09E-20 | 0.013066572 | Reject |
| 5 | 0.145259924 | 1.63E-27 | 0.013310899 | Reject |
| 6 | 0.101202199 | 1.56E-13 | 0.013682444 | Reject |
| 7 | 0.119321323 | 2.73E-17 | 0.014069671 | Reject |
| 8 | 0.100935386 | 1.93E-12 | 0.014310044 | Reject |
| 9 | 0.122428084 | 4.80E-17 | 0.014546554 | Reject |
| 10 | 0.120962512 | 4.48E-16 | 0.014845095 | Reject |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Probabilities | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | 0.021012179 | 2.44E-07 | 0.004067252 | Reject |
| 1 | 0.019957708 | 9.35E-09 | 0.003472383 | Reject |
| 2 | 0.025132635 | 4.22E-16 | 0.003084442 | Reject |
| 3 | 0.024546679 | 7.21E-16 | 0.003036695 | Reject |
| 4 | 0.031309101 | 3.71E-25 | 0.003011038 | Reject |
| 5 | 0.041423236 | 1.14E-35 | 0.003305109 | Reject |
| 6 | 0.030580739 | 5.01E-20 | 0.003326926 | Reject |
| 7 | 0.035748835 | 1.83E-22 | 0.003653008 | Reject |
| 8 | 0.036869475 | 3.75E-17 | 0.004366176 | Reject |
| 9 | 0.047995434 | 1.17E-18 | 0.005425359 | Reject |
| 10 | 0.051314495 | 2.95E-15 | 0.006483565 | Reject |

**Table 8.** Results including Slope, Standard Error, P-value, and Conclusion of regression slop test performed for Northeast region and the set of actual survival status, predicted survival Status, and predicted survival probabilities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Actual Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -1.31E-02 | 5.07E-02 | 6.70E-03 | Accept |
| 1 | -3.85E-02 | 8.69E-05 | 9.81E-03 | Accept |
| 2 | -3.27E-02 | 3.35E-03 | 1.12E-02 | Reject |
| 3 | -3.23E-02 | 7.27E-03 | 1.20E-02 | Reject |
| 4 | -1.94E-02 | 1.32E-01 | 1.29E-02 | Accept |
| 5 | -2.76E-02 | 4.14E-02 | 1.36E-02 | Reject |
| 6 | -2.00E-02 | 1.58E-01 | 1.42E-02 | Accept |
| 7 | -2.26E-02 | 1.20E-01 | 1.46E-02 | Accept |
| 8 | -2.25E-02 | 1.32E-01 | 1.49E-02 | Accept |
| 9 | -2.03E-02 | 1.84E-01 | 1.53E-02 | Accept |
| 10 | -1.96E-02 | 2.05E-01 | 1.54E-02 | Accept |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -1.19E-01 | 7.61E-23 | 1.21E-02 | Reject |
| 1 | -9.37E-02 | 7.21E-14 | 1.25E-02 | Reject |
| 2 | -4.53E-02 | 4.61E-04 | 1.29E-02 | Reject |
| 3 | -5.03E-02 | 1.48E-04 | 1.32E-02 | Reject |
| 4 | -5.14E-02 | 1.47E-04 | 1.35E-02 | Reject |
| 5 | -9.82E-02 | 1.55E-12 | 1.39E-02 | Reject |
| 6 | -4.76E-02 | 8.16E-04 | 1.42E-02 | Reject |
| 7 | -3.09E-02 | 3.31E-02 | 1.45E-02 | Reject |
| 8 | -3.04E-02 | 3.95E-02 | 1.47E-02 | Reject |
| 9 | -8.42E-02 | 2.27E-08 | 1.50E-02 | Reject |
| 10 | -6.97E-02 | 5.58E-06 | 1.53E-02 | Reject |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Probabilities | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -4.96E-02 | 1.64E-32 | 4.16E-03 | Reject |
| 1 | -3.43E-02 | 1.11E-21 | 0.003581 | Reject |
| 2 | -1.45E-02 | 6.62E-06 | 3.22E-03 | Reject |
| 3 | -1.69E-02 | 7.23E-08 | 3.13E-03 | Reject |
| 4 | -1.64E-02 | 1.55E-07 | 3.12E-03 | Reject |
| 5 | -3.08E-02 | 4.65E-19 | 3.44E-03 | Reject |
| 6 | -1.84E-02 | 1.13E-07 | 3.46E-03 | Reject |
| 7 | -1.20E-02 | 1.49E-03 | 3.77E-03 | Reject |
| 8 | -1.71E-02 | 1.43E-04 | 4.50E-03 | Reject |
| 9 | -3.63E-02 | 1.06E-10 | 5.61E-03 | Reject |
| 10 | -3.66E-02 | 4.46E-08 | 6.68E-03 | Reject |

**Table 9.** Results including Slope, Standard Error, P-value, and Conclusion of regression slop test performed for Southeast region and the set of actual survival status, predicted survival Status, and predicted survival probabilities.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Actual Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -2.74E-03 | 6.43E-01 | 5.90E-03 | Accept |
| 1 | -0.00921443 | 0.287999931 | 0.008671734 | Accept |
| 2 | -0.00778704 | 0.427971123 | 0.009823363 | Accept |
| 3 | -0.00765591 | 0.470636974 | 0.010611449 | Accept |
| 4 | -0.0209397 | 0.065237564 | 0.011356279 | Accept |
| 5 | -0.01105068 | 0.354125252 | 0.011925079 | Accept |
| 6 | -0.02409774 | 0.054279037 | 0.012518771 | Accept |
| 7 | -0.00837090 | 0.516714655 | 0.012909039 | Accept |
| 8 | -0.00627943 | 0.635475759 | 0.013246156 | Accept |
| 9 | -0.01317932 | 0.328455368 | 0.013485321 | Accept |
| 10 | -0.01504229 | 0.270069929 | 0.013637537 | Accept |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Status | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -1.45E-02 | 1.76E-01 | 1.07E-02 | Accept |
| 1 | -0.02803392 | 0.011354747 | 0.01107138 | Reject |
| 2 | -0.08360712 | 1.81E-13 | 0.012913625 | Reject |
| 3 | -0.09058845 | 8.00E-15 | 0.011640304 | Reject |
| 4 | -0.12751155 | 1.19E-26 | 0.011888209 | Reject |
| 5 | -0.12472677 | 1.40E-24 | 0.012143389 | Reject |
| 6 | -0.10407992 | 1.02E-16 | 0.012504388 | Reject |
| 7 | -0.12768173 | 1.96E-23 | 0.012751611 | Reject |
| 8 | -0.10599844 | 4.65E-16 | 0.013019163 | Reject |
| 9 | -0.08310523 | 4.11E-10 | 0.013275335 | Reject |
| 10 | -0.08427282 | 4.90E-10 | 0.013520699 | Reject |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predicted Survival Probabilities | | | | |
| Year | Slope | P-value | Standard Error | Conclusion |
| 0 | -6.04E-03 | 1.02E-01 | 0.003696503 | Accept |
| 1 | -0.01147718 | 0.000306026 | 0.00317791 | Reject |
| 2 | -0.02560045 | 1.35E-19 | 0.002819845 | Reject |
| 3 | -0.02509226 | 9.74E-20 | 0.002752753 | Reject |
| 4 | -0.03390140 | 6.44E-35 | 0.002736397 | Reject |
| 5 | -0.03617921 | 7.28E-33 | 0.0030151 | Reject |
| 6 | -0.02974223 | 1.92E-22 | 0.00304123 | Reject |
| 7 | -0.03727764 | 3.73E-29 | 0.003310149 | Reject |
| 8 | -0.0340464 | 1.39E-17 | 0.003976836 | Reject |
| 9 | -0.02777390 | 2.21E-08 | 0.004957406 | Reject |
| 10 | -0.02714015 | 4.51E-06 | 0.005912222 | Reject |

### Result of ANOVA

The results of ANOVA determine whether the patients from three different regions - middle west, north west, and north east – have same mean values of actual survival status, predicted survival status, and predicted survival probability.

As Table 10 suggested, the hypothesis is accepted in year 0, 5, 6, 7, 8, 9, and rejected in year 1, 2, 3, 4, which implies that there is no significant difference between the actual survival status of patients from different regions in year 0, 5, 6, 7, 8, 9. In the other years, the same result cannot be concluded.

For predicted survival status and predicted survival probability, the hypothesis is rejected for all the years, which suggests that there exist significant differences between the predicted survival status of patients from different regions, and also between the predicted survival probability of patients from different regions.

**Table 10.** Results including statistic, p-value, and conclusion of ANOVA performed for actual survival status of three regions.

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Survival Status | | | |
| Year | statistic | p-value | conclusion |
| 0 | 0.576839036 | 0.561695002 | Accept |
| 1 | 4.252878352 | 0.014277059 | Reject |
| 2 | 3.886647697 | 0.020558227 | Reject |
| 3 | 3.046513543 | 0.047590982 | Reject |
| 4 | 3.071013217 | 0.046443805 | Reject |
| 5 | 2.648017499 | 0.070874053 | Accept |
| 6 | 2.553425685 | 0.077903756 | Accept |
| 7 | 2.042358654 | 0.129822055 | Accept |
| 8 | 2.794213227 | 0.061254991 | Accept |
| 9 | 1.651012699 | 0.191960831 | Accept |
| 10 | 1.492085417 | 0.225007815 | Accept |

**Table 11.** Results including statistic, p-value, and conclusion of ANOVA performed for predicted survival status of three regions.

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted Survival Status | | | |
| Year | statistic | p-value | conclusion |
| 0 | 42.93389242 | 2.85878E-19 | Reject |
| 1 | 23.49654826 | 7.00458E-11 | Reject |
| 2 | 31.68267397 | 2.00546E-14 | Reject |
| 3 | 40.23185477 | 4.2943E-18 | Reject |
| 4 | 72.41616085 | 8.07779E-32 | Reject |
| 5 | 93.52238769 | 1.00755E-40 | Reject |
| 6 | 44.8989143 | 4.4961E-20 | Reject |
| 7 | 59.53555269 | 2.65187E-26 | Reject |
| 8 | 40.39343963 | 3.91277E-18 | Reject |
| 9 | 49.93880958 | 3.36598E-22 | Reject |
| 10 | 44.63539772 | 6.21978E-20 | Reject |

**Table 11.** Results including statistic, p-value, and conclusion of ANOVA performed for predicted survival probability of three regions.

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted Survival Possibilities | | | |
| Year | statistic | p-value | conclusion |
| 0 | 58.41875519 | 6.57133E-26 | Reject |
| 1 | 36.53722545 | 1.78681E-16 | Reject |
| 2 | 55.5414519 | 1.16897E-24 | Reject |
| 3 | 57.39093571 | 1.94586E-25 | Reject |
| 4 | 95.55743489 | 1.31515E-41 | Reject |
| 5 | 129.3908445 | 9.61627E-56 | Reject |
| 6 | 69.33773114 | 1.76977E-30 | Reject |
| 7 | 78.31290703 | 2.95523E-34 | Reject |
| 8 | 54.35978876 | 4.3223E-24 | Reject |
| 9 | 55.5421123 | 1.39396E-24 | Reject |
| 10 | 42.37661191 | 5.71776E-19 | Reject |

## Existence of Bias

For the protected variable gender, there is a noticeable distinct between the results of actual value and the predicted value. In most years, there is neither a significant linear relationship between gender and actual survival status, nor a significant difference between the actual survival status means of protected groups. However, a significant linear relationship is exposed between gender and predicted values, and a significant difference between the predicted values means of protected groups is also indicated.

In this case, the existence of bias based on gender is discovered.

The similar conclusion can be constructed from the results of regions. In the regression slope test, nearly no significant relationship exists between region and actual survival status, but significant relationships exist between region and predicted values. In ANOVA, for most years, there is no significant difference among the actual survival status means of regions, but the same conclusion is denied for the means of predicted values.

The existence of bias based on region is therefore proved as well.

# References

Ana I Balsa, T. G. (2005). Testing for Statistical Discrimination in Health Care. *Health Service Research*, 227-252.

Benjamin A Goldstein, A. M. (January 2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, 198–208.

Jee, C. (2019, Oct 25). *A biased medical algorithm favored white people for health-care programs*. MIT technology review

Jenna Wiens, E. S. (2018). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, 149–153.

Milena A. Gianfrancesco, S. T. (November 2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med*, 1544-1547.

Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., & Jha, N. K. (2015). Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare. *IEEE Journal of Biomedical and Health Informatics*.

RUGGIERI, A. R. (2013). A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review,*, 1-52.

Sam Corbett-Davies, S. G. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning.

Shenoy, J. W. (21 August 2017). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, 149–153.

Žliobaitė, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 1060–1089.