Preventing Recidivism in State Prisons

Summary

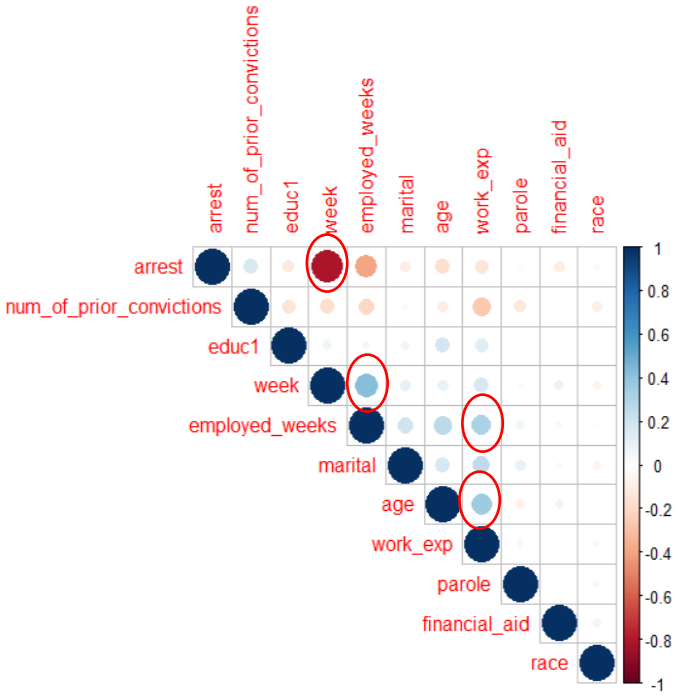
In evaluating the prisoner reform program based on the data of state prisoners, several interesting findings were uncovered. From the total 432 convicts which were released from state prisons, 50% of them were chosen at random to receive financial aid. Prisoners were then tracked, and additional metrics were observed and their impacts on prisoners getting re-arrested were measured in this study.

Methodology

The data cleaning step was conducted first. With the raw data obtained, categorical variables must be changed to dummy variables. This was done by converting education into a dummy variable by assigning 1s for prisoners who completed 12th grade or had some college. Those which did not complete college were assigned 0s. This data was stored in a new variable, educ1.

After this, further data cleansing efforts involved taking the sum of emp1 through emp52 variables to tally the number of weeks that each prisoner worked for. This data was stored in a new variable, employed_weeks.

With the data cleaned, a correlation plot was conducted to extract the variables which were highly correlated with one another. The correlation plot revealed the following:



The strongest correlations are found between week and arrest, week and employed_weeks, age and work_exp, and employed_weeks and work_exp. When thinking about the insights derived from the correlation plot, we can eliminate some less significant variables to further strengthen the model.

"Week" is removed as the week in which the prisoner committed a crime is not relevant since "arrest" tells us the same thing as week and is more relevant because we are primarily concerned with whether the prisoner was arrested rather than which week the arrest occurred. Additionally, "age" was removed as it conveyed similar information to other variables such as the amount of work experience a prisoner has, the marital status of the prisoner, and the number of prior convictions.

After this, each variable was independently run against "arrest" to determine which variables had a p-value significantly greater than 0.05 as these would also get removed from the final model. Doing this, both, "parole" and "race", had p-values which were significantly higher than 0.05 so these two variables were removed before running the logistic regression model.

Using arrest as a dependent variable and the rest of the variables as independent variables, running the logistic regression model gave us the following p-values of the remaining variables:

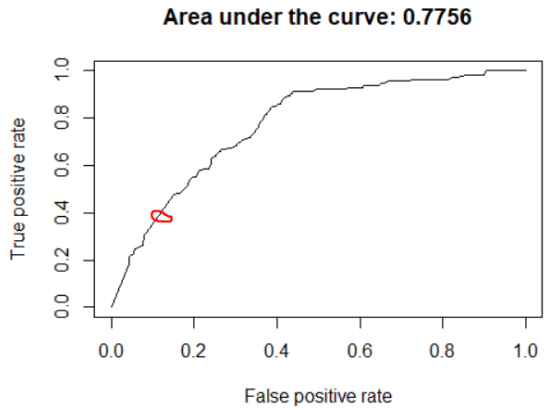
```
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.091487   0.282169   0.324   0.7458
financial_aid   -0.405708   0.243439  -1.667   0.0956 .
work_exp         0.070732   0.261483   0.271   0.7868
```

```
marital                   -0.152125   0.453894  -0.335   0.7375
num_of_prior_convictions   0.059318   0.040958   1.448   0.1475
educ1                     -0.962817   0.480400  -2.004   0.0450 *
employed_weeks            -0.062353   0.009127  -6.832 8.38e-12 ***
```

From this, we can see that only education and number of employed weeks have a p-value lower than 0.05 which indicates that these two variables can reliably predict whether arrests will be made for the prisoners. Although the rest of the variables have low p-values when individually run against arrest, a logistic regression with all these variables together displays that they are not significant when in conjunction with other variables due to them being correlated with one another. Additionally, financial_aid is somewhat significant as well at a p-value of just under 10%. After putting only these three variables, financial_aid, employed_weeks, and educ1, into a new logistic regression model, we can move forward with the impact these three variables have on arrests made. The co-efficient for education is -1.04 which indicates that prisoners who have at least a grade 12 education are less likely to be arrested again. Employed_weeks has a co-efficient of -0.06 which indicates the same effect as education but to a lesser intensity for convicts who found employment. Lastly, the coefficient of financial aid is -0.41 which indicates that financial aid is the biggest determinant of whether a prisoner will be arrested upon release. Not giving financial aid severely hampers the prisoner's chances of staying out of prison and can lead to them committing another crime.

Lastly, using these two variables in a logistic regression, we can run a confusion matrix which garners the following results:

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 250  48
         1  68  66

               Accuracy : 0.7315
                 95% CI : (0.687, 0.7727)
    No Information Rate : 0.7361
    P-Value [Acc > NIR] : 0.61057

                  Kappa : 0.3457

 Mcnemar's Test P-Value : 0.07771

            Sensitivity : 0.7862
            Specificity : 0.5789
         Pos Pred Value : 0.8389
         Neg Pred Value : 0.4925
```



These results indicate that we can use these two variables of education and weeks employed to derive a model with a 73% overall predictive power, 78% predictive power of which prisoners should be released from prison as they will not get arrested again and 57% predictive power of which prisoners ought to not be released from prison as they are likely to be arrested.

Recommendations

We can use this model by observing closely the prior education of prisoners and the weeks of experience the prisoners accumulate before deciding on whether the prisoners should be sent back to prison or not. All prisoners should be given financial aid since it lowers their chances of being re-arrested by 40% based on the coefficient. Additionally, using the findings of other variables having an impact on arrests made when individually observed, we should only consider prisoners who have at least completed their high school education, have prior work experience, have been married and have no prior convictions to be enrolled in a program where they are eligible to receive financial aid. After receiving the financial aid for a few weeks, prisoners should be observed carefully to determine whether they can find employment or not. If they cannot find employment after the financial aid has been given to them for a period up to 6 months, they should be re-arrested and sent to prison as they are likely to commit a crime based on this predictive model. If employment is found by the convicts within half a year then according to this model, those convicts have a high chance of remaining clean and not being arrested again. This step-by-step program will ensure the successful transition of former convicts into civilian life.