Evaluation of the Influence of Alcohol and Drugs on Parolees' Recidivism

September 7, 2019

Capstone Report for the Data Analysis and Interpretation Specialization

Introduction

The purpose of this study was to evaluate whether drug related factors (such as age of first marijuana usage and opiate usage in the past thirty days) or alcohol related factors (such as number of can beer drank per day) more heavily affect whether parolees commit crimes after 9 months of release.
Lots of TV shows based on crime investigation feature drug addicts much more prominently, and show that drug addiction comes back even after being released to the outside world. However, there isn't as much spotlight shone on alcohol addiction, and I want to see whether alcohol addiction really doesn't have much of an impact on parolees committing crimes after 9 months.
This would allow for the government to allocate its resource effectively so that they can better track parolees who are more prone to committing crimes, as well as putting greater regulation on alcohol or drug usage.

Methods

Sample

The sample included released parolees with a minimum of 3 months of parole (n=476) who were required as part of their parole to participate in a 12 week drug addiction treatment program. Parolees were interviewed before starting the program, 3 months, and 6 months after completing the program (9-month follow up).

Measures

The response variable was a categorical variable ANYCRIME that measured whether the parolee committed a crime or not.
The explanatory variables were loaded from the Intake Codebook, and included the following for the alcohol-related variables:
AGE1ALC - quantitative variable measuring age of first alcohol usage
ALC30D - categorical variable measuring the degree of alcohol usage for the past 30 days
ALCDYS - quantitative variable measuring number of days of alcohol usage for past 30 days
ALCPRB1 - categorical variable measuring the degree of alcohol affecting physical health for past six months
ALCPRB3 - categorical variable measuring the degree of alcohol affecting emotional health for past six months

OUNCE_LIQ – quantitative variable measuring number of glasses of alcohol consumed per day; created by summing the product of number of glasses by size of glass for each type of liquor

The following for the drug-related variables:

AGE1MJ - quantitative variable measuring age of first marijuana usage

MJ30D - categorical variable measuring the degree of marijuana usage for the past 30 days

AGE1COC - quantitative variable measuring age of first cocaine usage

COC30D - categorical variable measuring the degree of cocaine usage for the past 30 days

DRGPRB1 - categorical variable measuring the degree of drugs affecting physical health for past six months

DRGPRB3 - categorical variable measuring the degree of drugs affecting emotional health for past six months

Since dropping blank data points would result in only eight full samples of data being analyzed, blank data points were all converted to 0. 0 denies the presence of the variable in both categorical (ex. 0 in DRGPRB1 means that the parolee felt that they had no influence of drug on their physical health) and quantitative (ex. 0 in AGE1COC means that the parolee never did cocaine) explanatory variables, carrying the same meaning as having null data points

Analysis:

For univariate analysis, all variables' distributions were examined via frequency tables and graphed with the y-axis as the proportion of parolees who committed crimes after release.

For bivariate analysis, chi squared tests were used to examine each explanatory variable's relationship with the response variable. Other diagrams were employed to visualize the relationship if necessary.

Logistic regression was conducted to see which explanatory variables contributed to parolees having greater odds of committing crimes after release.

A random forest was created to evaluate the importance of the explanatory variables in predicting the response variable.

Univariate Analysis:

The following is a visual representation of the univariate analysis of the thirteen variables. Approximately 31% of the subjects committed crimes after being released; both alcohol and drugs were first consumed at adolescent ages; alcohol was much more frequently consumed but had less impact on health, and vice versa for drugs.

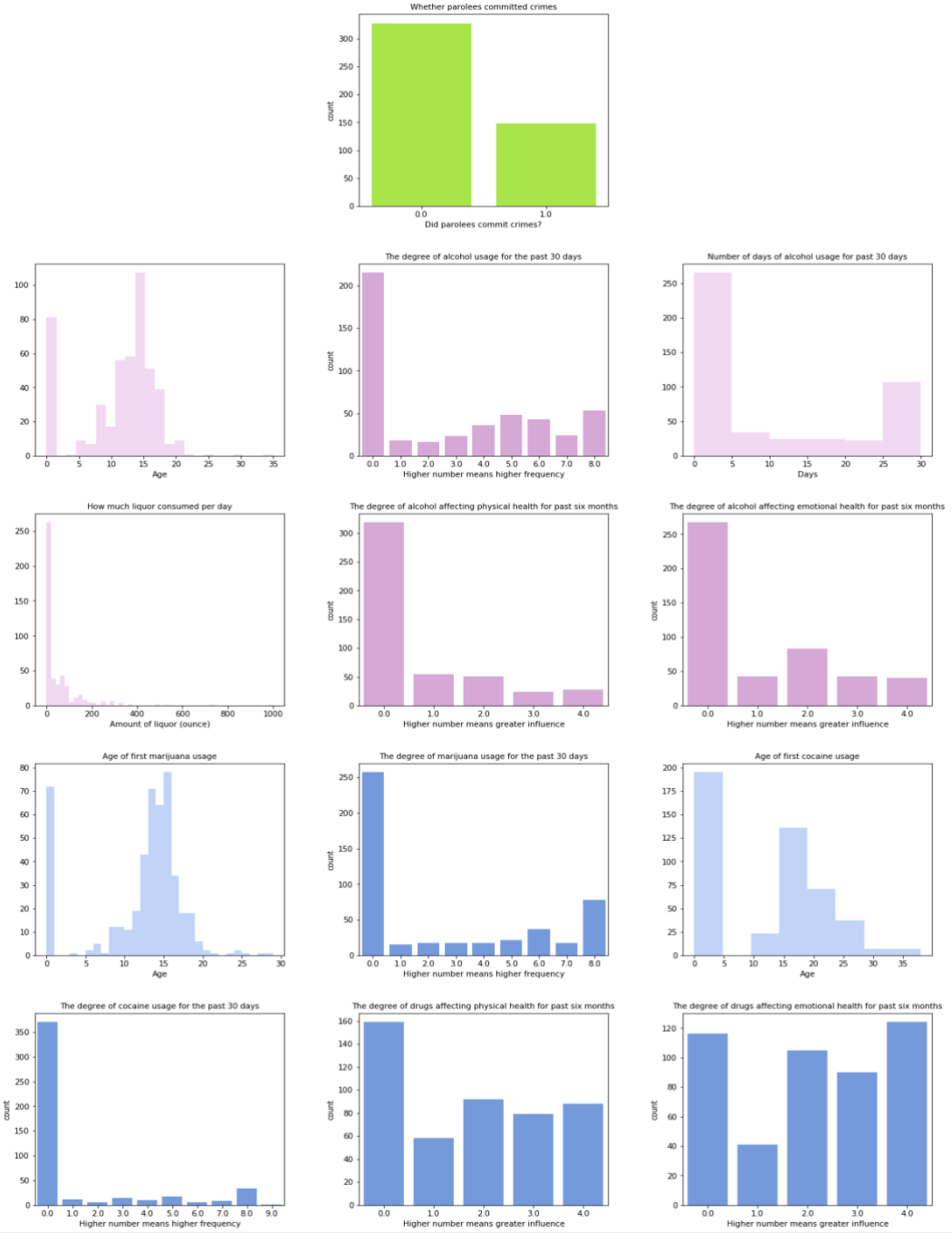

Figure 1. Frequency distribution of each of the explanatory variables as well as the response variable. Green is response variable; purple for alcohol related variables, blue for drug related variables

Bivariate Analysis:

Chi squared tests evaluating whether there is an association between the explanatory variables and the response variable were conducted for each of the twelve predictor variables.

Before conducting chi squared tests, quantitative explanatory variables were converted into categorical variables with five bins.

AGE1MJ, ALCPRB3, COC30D, DRGPRB3, and DRGPRB1 were (in that order) the variables with the most significant results in the chi squared tests. Four of those five are drug-related variables. All the variables showed a positive association with the response variable, and this makes sense because alcohol and drug prone people are likely to commit crimes.

The following is a table of chi squared values and the corresponding p values for each chi squared test:

	Chi squared values	P values
AGE1ALC	2.422129752916643	0.6586317957446013
ALC30D	4.072139407479838	0.8505568063583866
ALCDYS	3.9628740139551284	0.41105362210676144
OUNCE_LIQ	5.872566453980491	0.20886936039980927
ALCPRB1	4.711281448896497	0.3182245554266131
ALCPRB3	8.873495439697638	0.06434045512947154
AGE1MJ	14.117834242224486	0.006928319955472779
MJ30D	6.142209502707901	0.6313058729508346
AGE1COC	2.575579402103792	0.6311551635531009
COC30D	15.936989206501403	0.06820913434506487
DRGPRB1	6.928085673630688	0.13973772367297893
DRGPRB3	7.019721540568784	0.1348496953494122

Table 1. Chi squared values and p values for each explanatory variable

The following figure is a sample of some of the visualizations of the relationship between the explanatory variables and the response variable, with the vertical axis representing the proportion of parolees who committed crimes after being released.

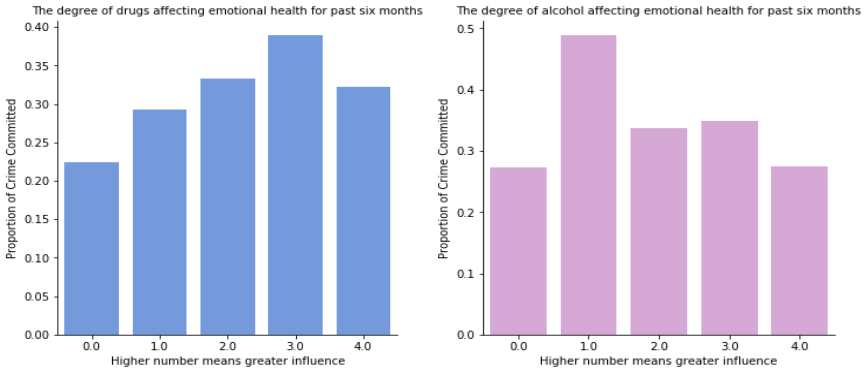

Figure 2. bar graphs of ALCPRB3 and DRGPRB3 vs ANYCRIME; these are the variables that had some of the most significant associations with the response variable.

Logistic Regression:

Amongst categorical variables, the odd ratios for DRGPRB3 (2.498205), COC30D (4.931632), ALC30D (1.879638), and ALCPRB3 (1.943097) were some of the highest – odds ratios on the lower end weren't analyzed because negative associations don't make sense unless the variable is age, and they weren't significant, as shown below. None of the odds ratios for quantitative variables were significant enough to be recorded. Although a great majority of the results were proven to be inconclusive due to confidence intervals of odds ratios containing 1, an approximately equal distribution of noticeably big odds ratios were present in both alcohol and drug related variables.

	Lower CI	Upper CI	OR
Intercept	0.131877	0.504191	0.257859
ALC30D[T.1.0]	0.481489	4.733820	1.509729
ALC30D[T.2.0]	0.387587	4.219224	1.278795
ALC30D[T.3.0]	0.244059	2.447274	0.772838
ALC30D[T.4.0]	0.424948	2.671469	1.065475
ALC30D[T.5.0]	0.743681	4.750751	1.879638
ALC30D[T.6.0]	0.471612	4.228802	1.412216
ALC30D[T.7.0]	0.209490	3.488553	0.854878
ALC30D[T.8.0]	0.510134	5.133989	1.618340
ALCPRB1[T.1.0]	0.512279	3.040784	1.248091
ALCPRB1[T.2.0]	0.396439	2.618443	1.018848
ALCPRB1[T.3.0]	0.362686	4.157015	1.227881
ALCPRB1[T.4.0]	0.207145	3.064422	0.796730
ALCPRB3[T.1.0]	0.800890	4.714284	1.943097
ALCPRB3[T.2.0]	0.495570	2.583644	1.131537
ALCPRB3[T.3.0]	0.350416	2.638552	0.961557
ALCPRB3[T.4.0]	0.210175	2.710841	0.754818
MJ30D[T.1.0]	0.097562	1.637074	0.399645
MJ30D[T.2.0]	0.250195	2.805073	0.837743
MJ30D[T.3.0]	0.422393	4.143388	1.322928
MJ30D[T.4.0]	0.146251	1.985067	0.538812
MJ30D[T.5.0]	0.645049	5.063443	1.807255
MJ30D[T.6.0]	0.585032	3.205171	1.369353
MJ30D[T.7.0]	0.583680	5.555379	1.800713
MJ30D[T.8.0]	0.418552	1.465135	0.783094
COC30D[T.1.0]	0.314073	5.856186	1.356197
COC30D[T.2.0]	0.250545	11.757781	1.716348
COC30D[T.3.0]	0.606421	7.497330	2.132261
COC30D[T.4.0]	0.159678	4.722611	0.868388
COC30D[T.5.0]	0.011950	0.812623	0.098542
COC30D[T.6.0]	0.161256	8.239775	1.152697
COC30D[T.7.0]	1.085003	22.415598	4.931632
COC30D[T.8.0]	0.740267	3.894603	1.697954
COC30D[T.9.0]	0.000000	inf	0.000000
DRGPRB1[T.1.0]	0.173472	0.997177	0.415912
DRGPRB1[T.2.0]	0.371306	1.657284	0.784448
DRGPRB1[T.3.0]	0.486659	2.483358	1.099340
DRGPRB1[T.4.0]	0.250469	1.544873	0.622048
DRGPRB3[T.1.0]	0.591685	3.887144	1.516563
DRGPRB3[T.2.0]	0.849200	4.187577	1.885760
DRGPRB3[T.3.0]	1.056406	5.907795	2.498205
DRGPRB3[T.4.0]	0.978068	5.735651	2.368513
AGE1ALC	0.945943	1.034607	0.989282
ALCDYS	0.956144	1.025990	0.990452
OUNCE_LIQ	0.998860	1.004293	1.001573
AGE1MJ	0.968436	1.062971	1.014603
AGE1COC	0.962313	1.011420	0.986561

Figure 3. confidence intervals and odds ratios for the predictor variables

Random Forest:

According to the table below, the six explanatory variables deemed to be most important in determining the response variable were ALCPRB3, AGE1MJ, AGE1ALC, MJ30D, DRGPRB1, and DRGPRB3 – four of which are drug related.

Variable name	Importance
AGE1ALC	0.10476944
ALC30D	0.07864462
ALCDYS	0.0753424
OUNCE_LIQ	0.05898458
ALCPRB1	0.05064108
ALCPRB3	0.1359279
AGE1MJ	0.10596623
MJ30D	0.09461535
AGE1COC	0.05901889
COC30D	0.07014221
DRGPRB1	0.0842937
DRGPRB3	0.08165359

Table 2. Relative importance scores of the explanatory variables

The accuracy of the random forest was 65%, and the subsequent growing of multiple trees did not affect the accuracy score very much, as the score oscillated between 0.60 and 0.68, suggesting that the interpretation of a single tree would suffice.

The confusion matrix was [115, 13] [54, 9]. True negatives were the most found (115), and true results (124) were predicted more than false ones (67).

Conclusion

Overview

Multiple statistical approaches were employed to analyze whether alcohol or drug had greater influence over the recidivism of parolees after release of six months. Initial analysis was done through univariate analysis of each of the thirteen variables. Chi squared tests revealed that drug related variables had more significant associations with recidivism rates, and random forest analysis came to a similar conclusion, while logistic regression analysis proved to be rather inconclusive compared to the previous two approaches. Data management included binning quantitative variables into categories for chi squared tests and replacing null data points with 0, as well as combining smaller variables to created a quantitative variable for amount of liquor consumed.

Implications

Such results indicate that, although the prevalence of alcohol may be of concern to many around the world (not just in the United States), the actual crime rates, in this case recidivism of parolees, are more strongly related to drug usage, which may not be as prevalent across the globe but seems to have just as much, if not a stronger, impact on crime rates. Thus the government should continue to keep its focus on drug addicts and drug criminals to eradicate and suppress crime rates not only in parolees but also in ordinary citizens and criminals as well.

Limitations

The data management provides limitations for this study. Since there were too many null data points, converting them to 0 was the best choice that could be made, but obviously a fuller data set with no null data points would be better for data analysis. If all null data points were eradicated, only eight parolees would have been the subject of analysis, and that is not ideal for data analysis. Also, having more corresponding variables, such as age of first alcohol usage and age of first marijuana usage, would be better for direct comparison of the variables.

Future Directions

The paper provided attempts to analyze the possible predictor variables of determining recidivism of parolees in terms of alcohol and drug usage, but there are definitely future steps that could be taken. There are other external factors that affect criminal activities – financial status, domestic violence, and divorce are some of such factors. Evaluating those factors against alcohol and drug could also be a future direction of research.

Code can be found in: https://pastecode.xyz/view/da94bcb2