# Staying Out of Trouble: Logistic Regression on the Iowa Department of Corrections Recidivism Strategies

**Abstract**
Understanding what prisoner traits are associated with rearrest after release (known as recidivism) is vital for the successful and effective implementation of assistance programs aimed at preventing recidivism. Here, we perform a logistic regression to develop a predictive model of variables that determine whether a prisoner will recidivate. We hypothesize that inclusion in an Iowa Department of Corrections recidivism-reduction programs is a key factor in predicting prisoner recidivism. Our model confirmed this hypothesis within in-sample data and was applied to out-of-sample data to validate predictive power.

## Background and Significance

Recidivism occurs when a prisoner is released from jail, relapses back into criminal behavior, and is rearrested. Between 2005 to 2010, 67.8% of released prisoners were arrested for a new crime within three years, and 76.6% were arrested within five years (Durose et al., 2014). Recidivism contributes to the ongoing national issue of overcrowded prisons, leading to the endangerment of the lives of inmates and corrections officers (Samuels, Jr., 2013). Preventing recidivism is crucial in addressing a problem like this along with others such as reintroducing former criminals back into society and halting future criminal activity.

Equally as important is understanding the risk of recidivism for offenders who are serving probation while residing in the community. Utilizing the Iowa Violence and Victimization Instrument, Prell, et al.(2016) created a robust model with strong predictive power for charges of violence and victimization, but fair power for drug offenses. In addition to analyzing the impact of differing offenses on recidivism, it is possible to assess the effectiveness of rehabilitation treatments. While it is established that well-administered programs and regimens can be effective in reducing recidivism rates, it is less clear which criminal characteristics these types of programs best target (Peters et al., 2015).

The results reported in our study provide insight regarding the effectiveness of multiple variables in predicting the likelihood of recidivism. One of the considered factors in this analysis is whether the criminal was part of a “target population”: a group of paroled prisoners who were included in a recidivism-reduction program (RRP) conducted by the Iowa Department of Corrections (IDOC) that focused on reducing recidivism rates in parolees (IDOC, 2016). This allows us to assess whether inclusion in IDOC’s RRP, along with other factors (i.e. race, sex, previous offense), are influential in predicting the recidivism outcome for released prisoners.

We hypothesize inclusion in IDOC’s RRP will be highly influential in predicting whether the prisoner experiences recidivism within three years of initial release. By creating an effective model focused on identifying simple indicators of recidivism likelihood, programs such as the one operated by the IDOC can more accurately select viable candidates for program assistance.

## Methods

### Acquisition of the Recidivism Dataset

We used a publically available dataset provided by the Iowa Department of Corrections (IDOC) for Iowan offenders released from prison between 2013 and 2016 (IDOC, 2016). Some of the key variables described offender demographics, the crimes committed prior to recidivism, whether an offender recidivates, and the information detailing an offender’s recidivism charge. If there were key variables missing, the entire row was removed. After removing rows with missing key information, we were left with 16,498 rows and 17 variables, where each row represents one individual offender.

### Using Logistic Regression to Predict Recidivism

We created a logistic regression model to predict the probability of offender recidivism within three years of release from prison. To evaluate the predictive power of our model, we randomly selected a subset of 12,374 observations and constructed a model based on this subset, termed the training data. To determine the best predictors of recidivism, we employed stepwise variable selection technique and assessed all variables at the 0.05 significance level. We employed a drop-in-deviance test to determine if inclusion in an RRP was necessary in our model. Afterward, we evaluated the model on a testing dataset with 4,124 observations and determined its accuracy in successfully predicting offender recidivism.

## Results

Variable Selection in Logistic Regression

Using stepwise variable selection technique, we constructed a model of seven variables, termed Model A (see Appendix), for the recidivism dataset that included the following variables: sex of offender, age when released from prison squared, whether an offender committed a felony or misdemeanor, whether an offender was charged with a drug crime, whether an offender was charged with a public order crime, whether an offender was charged with a violent crime, and whether an offender was released on discharge or parole. Another model was created that included the seven variables in the first model and three other variables related to IDOC's RRP, termed Model B (see Appendix): whether an offender was included in IDOC's RRP, an interaction term of inclusion in the RRP with whether an offender was released on discharge or parole, and an interaction term of inclusion in the RRP with whether an offender committed a felony or misdemeanor.

Because we are particularly concerned with the effectiveness of IDOC's recidivism strategies, we conducted a drop-in-deviance test to determine whether inclusion in IDOC's RRP and the two interaction terms including this variable should be included in the model. Our analysis yielded a considerable score of $G = 139.77$ and p-value $p < 0.0001$ ($df = 3$). This allows us to reject the null hypothesis and conclude that there is evidence to support the inclusion of variables dealing with IDOC's RRP is necessary in our model. This means that Model B is both parsimonious and as powerful as Model A in predicting recidivism.

Evaluation of the Recidivism Logistic Regression Model

To determine the effectiveness in out-of-sample data, we applied our model on the testing dataset and observed how accurate the model would predict offender recidivism (see Appendix). When comparing our predicted outcomes based on Model B and the actual recidivism outcomes, we obtain a prediction accuracy of 68.94% using the training dataset and 68.70% using the testing dataset (see Appendix). We also produced a receiving operating characteristic (ROC) curve of Model B and the actual recidivism outcomes from the testing dataset to measure the model's classifier performance in correctly predicting recidivism. The ROC curve allows us to visualize the tradeoff between correctly predicting and incorrectly predicting offender recidivism. The AUC value of the ROC curve produced is 0.63 (Figure 1).

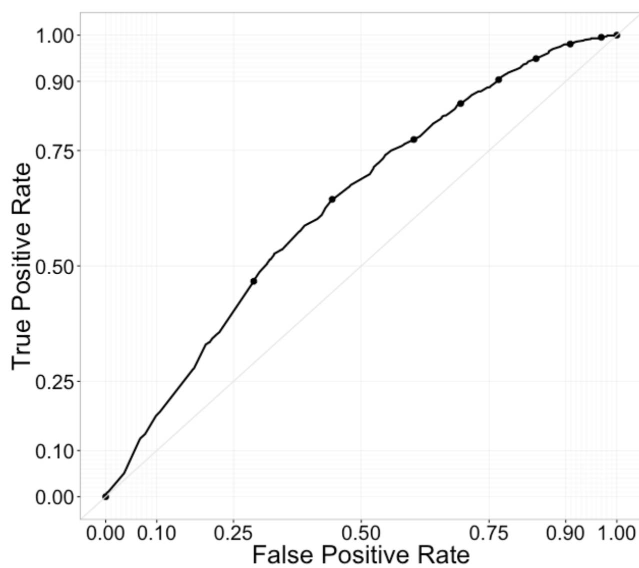


**Figure 1.** Receiving operating characteristic curve for the application of Model B onto the testing dataset. The curve displays the tradeoff between the rate at which it correctly predicts recidivism and the rate of incorrectly predicting recidivism (AUC = 0.63).

**Discussion**

The results from our drop-in-deviance test confirm our hypothesis that inclusion in the RRP of IDOC's recidivism-reduction programs implemented is effective in predicting whether a prisoner will be rearrested within three years of release. While this may seem to be resounding support of this program, it is important to remember the scope of this project is only analyzing IDOC's RRP and not different demographics or other regions' programs.

Model B is effective in predicting recidivism on in-sample data, but we need recognize the predictive power or the model as a whole. Applying our model to our testing dataset, we successfully predict offender recidivism 68.70% of the time (see Appendix). Although we obtain a substantially high accuracy, our model harshly overestimates the number of offenders who do not recidivate. Figure 1 maps the true positive rate against the false positive rate of our model, which displays its low predictive power, while the associated AUC value of 0.63 confirms that our model is not all that effective in predicting recidivism likelihood.

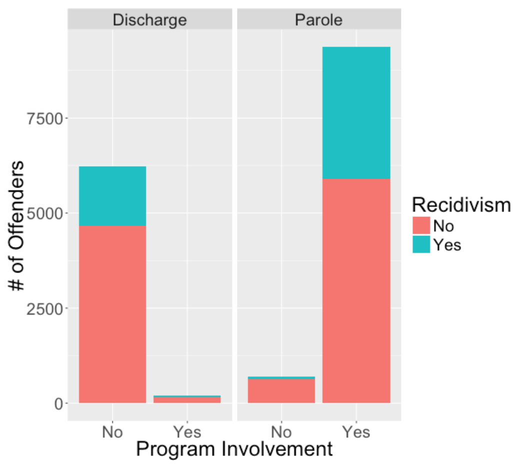


**Figure 2.** Number of offenders who were either discharged or released on parole that were involved in a RRP and recidivated.


**Figure 2.** Number of offenders who were either charged with felony or misdemeanor that were involved in a RRP and recidivated.

This may stem from multiple reasons. Our data relies heavily on the use of categorical variables with two levels (i.e. sex, offense classification, etc.). The use of binary variables reduces the amount of information each variable's output provides to the model, which may be limiting the applicable effectiveness of our final model. Furthermore, there is an unequal proportion of offenders involved in RRP programs. Figure 2 shows the proportion of paroled offenders in the RRP is higher than those not in the program. This trend is one which may stem from participant selection practices more than program effectiveness. Figure 3 shows the same trend with felony offenders. Therefore, the model may predict a RRP to have adverse effects on a prisoner's likelihood of recidivism based on other individual characteristics.

Developing effective models to determine recidivism has been a challenging task confronting many criminal justice researchers and practitioners. Despite the low predictive power in our model, other models assessing the likelihood of recidivism generate an accuracy of 0.63 to 0.64 with AUC values ranging from 0.60 to 0.70. These additional examples emphasize the difficulty of constructing such models (Ngo et al., 2014). Models that succeed in accurately predicting relapses into criminal behavior typically possess an AUC value above 0.80. Albeit the predictive power of our model was not as strong as initially thought, we do provide insight on recidivism strategies that may be impactful in preventing the recurrence of offender recidivism.

**References**


Iowa Department of Corrections (IDOC). 3-Year Recidivism for Offenders Released from Prison. Last updated September 23, 2016. Retrieved from https://data.iowa.gov/Public-Safety/3-Year-Recidivism-for-Offenders-Released-from-Pris/mw8r-vqy4

Durose, M.R., Cooper, A.D., Snyder, H.N.. Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010. Bureau of Justice Statistics (2014). NCJ 2444205.

Ngo, F.T., Govindu, R. & Agarwal, A. Assessing the Predictive Utility of Logistic Regression, Classification, and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct. American Journal of Criminal Justice (2015) 40: 47. doi:10.1007/s12103-014-9246-6

Peters, D.J., Hochstetler, A., DeLisi, M. Kuo, H.J. Parolee Recidivism and Successful Treatment Completion: Comparing Hazard Models Across Propensity Models. Journal of Quantitative Criminology (2015) 31: 149. doi:10.1007/s10940-014-9229-2

Prell, L., Vitacco, M. J., Zavodny, D. Predicting Violence and Recidivism in a Large Sample of Males on Probation or Parole. International Journal of Law and Psychiatry (2016) 49. 107-113.

Samuels, Jr., C. E. Statement of Charles E. Samuels, Jr. Director of the Federal Bureau of Prisons Before the U.S. Senate Committee on the Judiciary for a Hearing on the Oversight of the Federal Bureau of Prisons. Department of Justice (2013). Retrieved from https://www.justice.gov/iso/opa/ola/witness/11-06-13-bop-samuels-testimony-re-oversight-of-the-federal-bureau-of-prisons.201312231.pdf

Appendix

**Table 1. Logistic Regression Model (Model A)**

| Variable | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -0.119202149 | 0.086269157 | -1.381747 | 0.167049398 |
| Male | 0.404090252 | 0.062293462 | 6.486880612 | 8.76E-11 |
| ReleaseAge2 | -0.000240341 | 2.65E-05 | -9.081643513 | 1.07E-19 |
| Felony | -0.360065014 | 0.056124309 | -6.415491289 | 1.40E-10 |
| Drug | -0.154482194 | 0.048837023 | -3.163218895 | 0.00156035 |
| PublicOrder | -0.432769981 | 0.064795456 | -6.679017477 | 2.41E-11 |
| Violent | -0.640623613 | 0.058103315 | -11.02559483 | 2.88E-28 |
| Discharge | -0.523343432 | 0.044924157 | -11.64948798 | 2.31E-31 |

**Table 2. Logistic Regression Model (Model B)**

| Variable | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -1.464163765 | 0.167803128 | -8.725485521 | 2.65E-18 |
| Male | 0.411652082 | 0.062686734 | 6.566813397 | 5.14E-11 |
| ReleaseAge2 | -0.000250311 | 2.66E-05 | -9.400681209 | 5.42E-21 |
| Felony | -0.516600333 | 0.071088007 | -7.267053245 | 3.67E-13 |
| Drug | -0.158223265 | 0.049351745 | -3.206031843 | 0.001345791 |
| PublicOrder | -0.422521534 | 0.065140114 | -6.486349287 | 8.79E-11 |
| Violent | -0.643627549 | 0.058641827 | -10.97557121 | 5.01E-28 |
| Discharge | 0.935136701 | 0.145448118 | 6.429348945 | 1.28E-10 |
| TargetPop | 1.236413922 | 0.174479937 | 7.086281341 | 1.38E-12 |
| Discharge:TargetPop | -1.736456555 | 0.249136162 | -6.969909709 | 3.17E-12 |
| Felony:TargetPop | 0.370964678 | 0.112182206 | 3.306804987 | 0.000943666 |

**Table 3. Predicted recidivism outcomes based on Model B against actual outcomes in training data**

| | Predicted Recidivism | | |
|---|---|---|---|
| **Actual Recidivism** | No | Yes | Total |
| No | 8521 | 21 | 8542 |
| Yes | 3822 | 10 | 3832 |
| Total | 12343 | 31 | 12374 |

**Table 4. Predicted recidivism outcomes based on Model B against actual outcomes in testing data**

| | Predicted Recidivism | | |
|---|---|---|---|
| **Actual Recidivism** | No | Yes | Total |
| No | 2822 | 6 | 2828 |
| Yes | 1285 | 11 | 1296 |
| Total | 4107 | 17 | 4124 |