



Pandemic Era Policing: A Model of Felony Arrests in California

Richard Paul Yim
University of California, Davis

Introduction

Increase in Crimes? Various news outlets have been reporting an increase in crime across the nation [1,2,3]. California in particular represents a large portion of these crimes. With soft-on-crime policies naturally favored by the democratic-majority state government, and the Defund the Police Movement of recent years due to public outrage due to events such as the George Floyd incident, law enforcement agencies have developed poor public image.

Law Enforcement Shrinkage? Alarmingly, 2021 marks the first year where there has been a decrease in law enforcement personnel in police and sheriff departments across California, dropping to levels before 2017 [4]. With all of these recent trends a natural question is to ask how felony arrests are being made in recent years by law enforcement in California on the basis of social factors. We study this question with the generalized linear model.

Data

What does the data look like? The dataset consists of labels Y coded levels 1, 2 and 3, corresponding to low, medium and high counts of felony arrests. These labels are the results of aggregated counts based off of predictors of age, race and gender, as well as year of the felony arrest.

What about the design matrix? The corresponding design matrix for our study is shown below:

$$\mathbf{X} = [\mathbf{1} \quad \mathbf{1}_{\text{Male}} \quad \mathbf{1}_{\text{Hispanic}} \quad \mathbf{1}_{\text{Other}} \quad \mathbf{1}_{\text{White}} \quad \mathbf{1}_{20 \text{ to } 39} \quad \mathbf{1}_{40 \text{ and over}}],$$

where X has 144 rows and 7 columns with entries being one-hot encoded. The baseline for the gender indicator is female; the baseline for age group is 19 and under; and for race the baseline is Black.

Original data? Several preprocessing steps were taken on the original data:

1. The original data response consisted of different types of felonies. For this study, we combine all felonies into total felonies, and bin felonies uniformly it ordinal values of low, medium and high counts.
2. Age group was original split into six groupings, but is reduced to just three.
3. Location of county of the felony arrest was provided, but was aggregated in this study.

Research Question

Whodunnit? In the context of our study we are interested in the interaction between civilians and law enforcement in California in recent years. Naturally, we ask,

What are the likelihoods for propensities of felony arrests by individuals on the basis of race, gender and age in recent years?

We study this question by performing proportional odds model, multinomial regression.

Why proportional odds? Our response variable is ordinal in nature, and the proportional odds model provides natural interpretation of a *cumulative effect* as response levels increase. In other words, we are able to make statements such as, "the likelihood for the propensity of felony arrests is greater/lower for individuals..."

Methods

Generalized Linear Models: GLMs are powerful statistical models with useful asymptotic properties. They can be used to model expected response values of some dependent variable Y under a *link* function, g , through a linear relationship of predictors, X , or

$$g(\mathbb{E}(Y|\mathbf{X})) = \mathbf{X}\beta.$$

Multinomial Regression The multinomial regression model that is used the *proportional odds model*, with response categories coded as $j = 1, 2, \dots, M$ with a cumulative effect response $z_{im} = \sum_{j=1}^m y_{ij}$, which is equal to 1 for $j \leq m$, and zero otherwise. In particular, the GLM under the proportional odds model is

$$g(\mu_{im}) = \beta_{0m} + X_i\beta,$$

where we have that $\mu_{im} = E(z_{im})$, and we have ordered intercepts $\beta_{01} \leq \beta_{02} \leq \dots \leq \beta_{0,M-1}$, and $\beta \in \mathbb{R}^{p-1}$ a parameter vector acting on our design matrix X . In our study, our link function g is the logit link, our we have three classes with $M - 1 = 2$;

Stepwise BIC: We use the Bayesian Information Criterion (BIC) to produce the most parsimonious model possible. In particular, using stepwise BIC regression, we end up with a model corresponding to our design matrix \mathbf{X} specified in the Data section. The BIC of our resulting proportional odds model is 225.018.

Goodness of Fit

Model Validity: We can measure how well our model fits the data by observing the model's Pearson residuals by fitted values. Close to zero valued, constant residual relationships indicate good fit as seen in Figure 1.

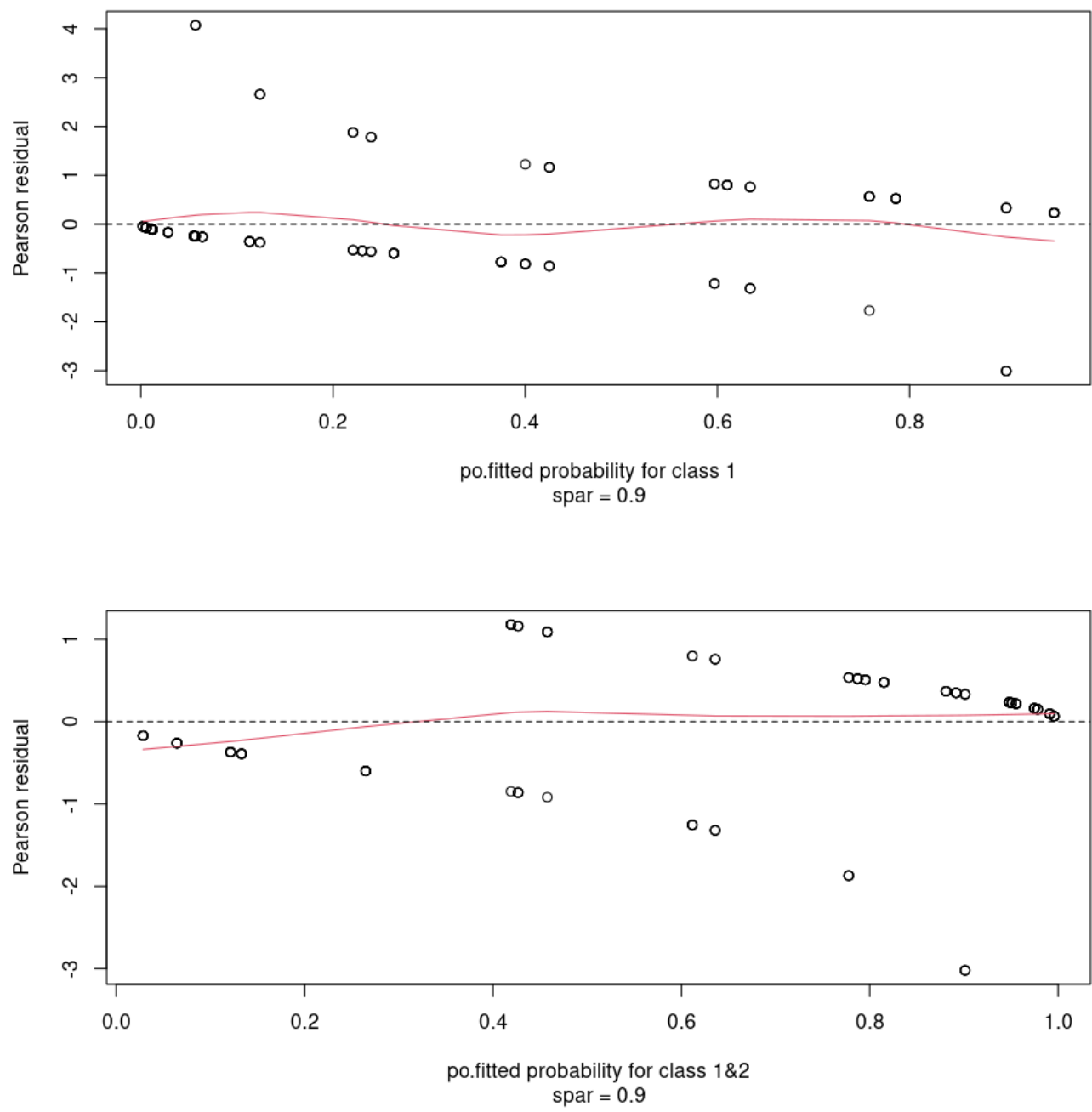


Figure 1. Goodness-of-fit plot for low propensity and low-medium propensity residual by fitted values.

Results

Table 1. 95% confidence intervals of β coefficients from proportional odds model.

		2.5 %	97.5 %
GENDER	Male	0.88	2.48
RACE	Hispanic	0.52	2.63
RACE	Other	-2.91	-0.78
RACE	White	-0.31	1.73
AGE_GROUP	20 to 39	2.94	5.14
AGE_GROUP	40 and over	-0.13	1.65

From our proportional odds model we are left with Table 1 providing 95% confidence intervals and conclude the following:

1. Men are **more** likely to have a greater propensity of felony arrests compared to females. This is evident by the 95% confidence interval range being greater than 0.
2. Hispanics are **more** likely to have a greater propensity of felony arrests compared to Blacks. This is evident by the 95% confidence interval range being greater than 0.
3. Other races are **less** likely to have a greater propensity of felony arrests compared to Blacks. This is evident by the 95% confidence interval range being less than 0.
4. Adults aged 20 to 39 are **more** likely to have a greater propensity of felony arrests compared to individuals ages 19 and under. This is evident by the 95% confidence interval range being greater than 0.

Conclusion

Limitations... The resulting conclusions are a bit limited. Generally, men would naturally have higher proclivity for acts of crime just by physiological conditioning and more reckless and questionable behaviors. But, we have some interesting insight with Hispanics having greater likelihoods that Blacks for felony arrests.

Economic Data... Data is lacking with respect to individuals' economic backgrounds. Future studies could perhaps survey and incorporate data such as parental education and/or occupation, nonintact family statuses, exposure to childhood violence abuse and more.

County Data... A missed opportunity of this data is applying median income, or other economic measure at the county level historically.

Acknowledgements

Thanks to Han and Professor Müller for the teaching. I learned a lot! Thank you UC Davis Math Department for helping me out with tuition ;).

References

1. <https://slate.com/business/2023/03/rising-crime-broken-windows-policing-revisited.html>
2. <https://www.oregonlive.com/news/2023/03/alarming-rise-in-hate-crimes-in-us-oregon-in-2021-fbi-stats-show.html>
3. <https://www.planetizen.com/blogs/122113-red-cities-blue-cities-and-crime>
4. <https://data-openjustice.doj.ca.gov/sites/default/files/dataset/2022-08/>