

The U.S. family firearm hold rate

Zikun Lei

Due Date: 12/22/2020

Code and data supporting this analysis is available at:
<https://github.com/leizikun/The-U.S.-family-firearm-hold-rate.git>

I am interested in finding out the family firearm hold rate in the U.S. In this report, I will be using the post-stratification analysis based on a logistic regression model. The results shows that around 37.5%, which is more than one third families in the U.S. hold firearms in their house.

Key words: gun, U.S., firearm, self defense, cartridge.

Introduction

Recently, because of the COVID-19 pandemic and the 2020 president election, many protesters are coming out and join the demonstration. Among those protesters, some of them are equipped with firearms which is extremely dangerous. The crime rate is experiencing an upward trend. Meanwhile, some news shows that residents in the U.S. store firearms and cartridges in their house to protect themselves. Meanwhile, because of the castle law in the U.S., I believe that more and more families are storing guns for self-defense. Hence, I would like to see the family firearm hold rate in the U.S.

I will use `survey_data` sample from UCLA Nationscape released in September 2020 to do the logistic regression and `census_data` sample from IPUMS released in 2018 to do the post-stratification. Since whether hold guns or not is a boolean variable, we will be using logistic regression model. Such kinds of variables fit well with the logistic regression model. This model will be run on R.studio.

Model

Model Specifics

From the survey dataset, five factors, age, sex, `census_region`, employment, `race_ethnicity` are chosen as the predictor variables. I choose sex as one factor from the data since I believe sex is an important factor that affect the rate of gun holds. In a recent report by Horowitz in 2020 August, it shows that men are more likely to have guns than women do. Furthermore, Horowitz claimed that “Women who own guns tend to become gun owners at a later age than men (Horowitz, 2020).” And that is the reason I put age as another factor. `census_region` is also chose because firearm law is slightly different in different region in the U.S. and status of employment will affect one’s ability to purchase firearms since the price of guns and cartridges are not cheap. Lastly, from my perspective, people’s race will be an potential lurking variable. The response variable is `household_gun_owner`. It shows 1 if the respondent’s family does hold guns at their house and otherwise it shows 0.

The subscripts of X are the names of those predictor variables. Every β represents the corresponding feature’s slope. β_0 is the intercept and this represents the general rate that people hold guns at home without

looking at other features. β_1 and β_2 are the slopes of the respondent's age and sex. β_3 to β_5 are different slopes of respondent's region. β_6 to β_7 represents their employment status' slope. β_8 to β_{11} represents different races.

And below is our model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_{age} + \beta_2 X_{sex: male} + \beta_3 X_{census_region: Northeast} + \beta_4 X_{census_region: South} + \beta_5 X_{census_region: West} + \beta_6 X_{employment: not in labor force} + \beta_7 X_{employment: unemployed} + \beta_8 X_{race_ethnicity: Asian or pacific islander} + \beta_9 X_{race_ethnicity: Black, or African American} + \beta_{10} X_{race_ethnicity: Some other race or multiraces} + \beta_{11} X_{race_ethnicity: White}$$

Post-Stratification

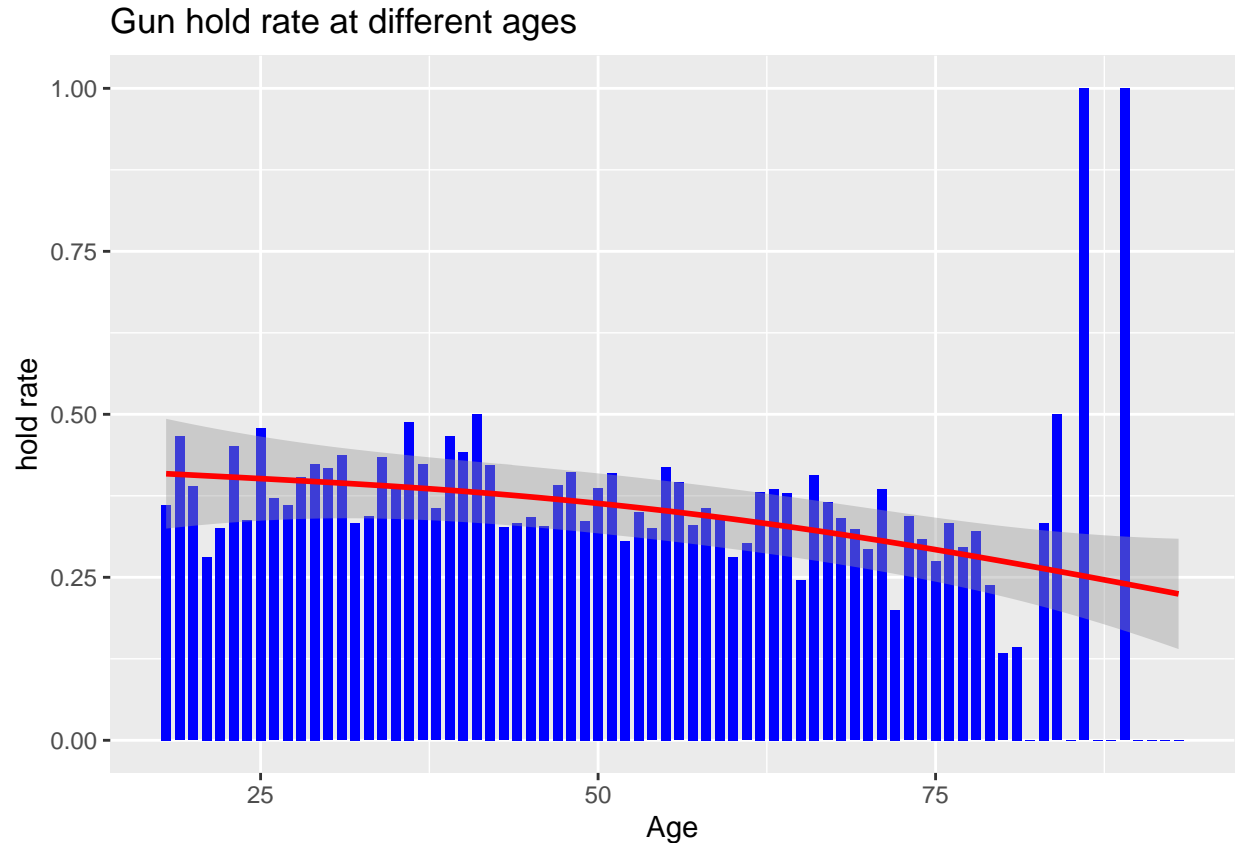
"Post-stratification is widely recognized as an effective method for obtaining more accurate estimates of population quantities in the context of survey sampling (Little, 1993)". This technique is applied since it can improve the accuracy estimates by increasing precision and reducing bias. I split the selected census_data into cells by age instead of splitting by other factors because I want to split census_data into smaller cells so that more variables can be presented.

Results

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.4095258	0.2395050	1.709885	0.0872872
age	-0.0115106	0.0018383	-6.261558	0.0000000
as.factor(sex)male	0.3775932	0.0558425	6.761751	0.0000000
as.factor(census_region)Northeast	-0.5196007	0.0884173	-5.876684	0.0000000
as.factor(census_region)South	0.2456510	0.0733076	3.350960	0.0008053
as.factor(census_region)West	-0.1769494	0.0839067	-2.108884	0.0349546
as.factor(employment)not in labor force	-0.1658695	0.0739066	-2.244312	0.0248123
as.factor(employment)unemployed	-0.3317744	0.0713073	-4.652739	0.0000033
as.factor(race_ethnicity)Asian or pacific islander	-1.3498770	0.2635878	-5.121166	0.0000003
as.factor(race_ethnicity)Black, or African American	-0.8866740	0.2359710	-3.757555	0.0001716
as.factor(race_ethnicity)Some other race or multiraces	-0.7721057	0.2454397	-3.145806	0.0016563
as.factor(race_ethnicity)White	-0.3081497	0.2235978	-1.378142	0.1681593

Figure 1: Key summaries of the logistic regression model for gun holds.

Figure 1 shows the outcome of the model we applied above. One variables does seem to have significant p-values since it is larger than 0.05 which is the variable of white people in race_ethnicity. This variable will have very little impact to the results. Other variables are believed to have a positive effect on the results since they have small p-values. As expected, male respondents are more likely to have guns in their house than female does. People whose race are Asian or pacific islander do not seem to be willing to hold guns which exerts a relatively large negative influence to the model result. Each category variable's first category is omitted as 0 and other categories are compared with it to see their impact to the first category. For example, in the sex category, female is omitted as 0 and male is used to compared with it. Based on the post-stratification result I calculated, around 37.5% families do hold guns in their house in the U.S.



Graph 2: Rate of gun holds in different ages.

Graph 2 above shows gun hold rates at respondent's house in different ages. We can see that ages of 85 and younger are gathered at a rate around 40% to 50%. Ages from 85 and above shows an surprising result with 100% gun hold rate. That means everyone's family whose age is older than 85 do hold guns in their house in this survey. Overall, the graph shows a downward trend with a very low slope value as age increases.

Discussion

Summary

The U.S. family firearm hold rate are calculated by using the 2018 IPUMS data and 2020 Nationscape data. I used a logistic regression model with a post-stratification to find out the family firearm hold rate with five features. Finally, the result shows around 37.5% families in the U.S. do hold guns in their house. We also find that male and people live in the south region has higher possibility to have guns in their house. an 100% gun hold rate are found for people of age 85 and above. However, this is due to lack of samples in this age group. From the overall trend, we can see a smooth downward line as age increases.

Conclusion

Since around 37.5% of the respondents hold guns in their family, we can see that more than one third families already own guns. The result in figure 1 shows that white people in the race category may not have enough influence to affect family firearm hold rate. Residents who live in the south region and male respondents have positive influence to the model. Compared to this, people who lives in other region, unemployed, not in labor force, and races exert a negative influence to the model. This indicates that people who have those features are less likely to purchase guns and store them at home.

Weaknesses

As we discussed in the beginning abstract, the U.S. is experiencing an upward trend in crime rates. Data we used are from 2018 and 2020 September which were collected before the U.S. federal election, data is outdated and results are smaller than what is actually shall be since other reports claim that more people are purchasing guns and keep them at home in this period than usual. Also, since category variables' name and outputs are slightly different of `survey_data` and `census_data`, I will have to filter some factors and recategorize them. For example, "gender" is collected in `survey_data` and "sex" is collected in `census_data`. Usually, gender contains "male", "female", and "other" while sex only contains "male" and "female". When recategorizing those into one same category, proportion outcome may be affected and results may be affected

Next Steps

Household income is considered to be an upgraded variable to replace employment status since that might reflect better financial status on firearm purchasing. Hence, a similar analysis will be conducted by replacing this variable and new data collected in the period of COVID-19 pandemic and U.S. election will be used to get a more accurate and recent results. After finding out the new U.S. family firearm hold rate, a comparison between this report and the newer one will be created to see which age group are more likely to purchase guns during the turmoil period.

References

1. Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. (2020). IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS. Retrieved from <https://doi.org/10.18128/D010.V10.0>
2. Tausanovitch, Chris and Lynn Vavreck. 2020. Democracy Fund + UCLA Nationscape, October 10-17, 2019 (version 20200814). Retrieved from <https://www.voterstudygroup.org/publication/nationscape-data-set>
3. Castle Doctrine from State to State. (n.d.). Retrieved from [https://www.southuniversity.edu/news-and-blogs/2016/08/castle-doctrine-from-state-to-state-46514#:~:text=The Castle Doctrine \(also known,some instances even deadly force.](https://www.southuniversity.edu/news-and-blogs/2016/08/castle-doctrine-from-state-to-state-46514#:~:text=The Castle Doctrine (also known,some instances even deadly force.)
4. Pandemic and protests spur Americans to buy guns at record pace. (2020, July 31). Retrieved from <https://www.theguardian.com/us-news/2020/jul/31/americans-guns-coronavirus-protests>
5. Horowitz, J. M. (2020, August 18). How male and female gun owners in the U.S. compare. Retrieved from <https://www.pewresearch.org/fact-tank/2017/06/29/how-male-and-female-gun-owners-in-the-u-s-compare/>
6. Tidyr. (n.d.). Retrieved from https://www.rdocumentation.org/packages/tidyr/versions/0.8.3/topics/drop_na

7. kableExtra. (n.d.) Retrieved from <https://www.rdocumentation.org/packages/kableExtra/versions/1.2.1>
8. Grouped data. (n.d.). Retrieved from <https://dplyr.tidyverse.org/articles/grouping.html>
9. Smoothed conditional means - geom_smooth. (n.d.). Retrieved from https://ggplot2.tidyverse.org/reference/geom_smooth.html
10. Reilly, C., Gelman, A., & Katz, J. (2000, July). Poststratification Without Population Level Information on the Poststratifying Variable, With Application to Political Polling. Retrieved from <https://stat.columbia.edu/~gelman/research/published/aprvlRv1.pdf>