American Research Journal of Humanities and Social Sciences

Volume 8, Issue 1, 78-85 Pages
Review Article | Open Access
ISSN (Online)- 2378-7031
DOI : 10.21694/2378-7031.22012

AMERICAN RESEARCH JOURNALS
An Academic Publishing House

Gun Ownership in the United States: Development and Validation of a Predictive Model

Jiayu Wu
Baldwin School, PA, USA.

ABSTRACT

Background

Objective: *This study aims to develop and validate a predictivelogistic regression model of household ownership of firearm in the United States.*

Methods: *Data from the Behavioral Risk Factor Surveillance System2017 was used. The list of potential predictor variables is largely based on existing literature on factors associated with gun ownership.Stepwise logistic regression analysis was employed to build the model. The model was then tested using Kolmogorov-Smirnov (KS) statistic and Area under the ROC Curve (AUC) as metrics to measure if the model is a good fit.*

Results: *About 48.8% participants reported having any guns in their families. From stepwise logistic regression analysis, 12 variables out of 14 are selected in the final prediction model. Factors that affect the likelihood of gun ownership include income level, education, veteran status, marital status etc.The resulting model is promising, with 72% percent of accuracy according to the ROC and a KS of 0.35.*

Conclusion: *A predictive model of gun ownership among U.S. households was developed and validated.*

KEYWORDS: *gun ownership, predictive model, Logistic regression*

BACKGROUND

In this research project, a predictive model of gun ownership in the United States is developed and validated. With the model, it is easy to see resident and household characteristics that are associated with possessing a gun. It will be helpful inidentifying and providing educationof safe gun storageto these householdsachieved based on the model.

STUDY METHODS

Data Source

Data from the Behavioral Risk Factor Surveillance System (BRFSS) is used. BRFSS is a nation-wide health surveys initiated by the Centers for Disease Control and Prevention (CDC) in the year 1984. It collects information on U.S. residents' health risk behaviors, preventive health practices, and health care access. It has been a timely and accurate source of data on health-related behaviors for many states.

For this study, the most recent data collected was used: BRFSS 2017 data.

Development and validation of the prediction model

Overall, the prediction model that the research aims to develop and validate is a logistic regression model.Data was split into two random samples: a 75% training sample for developing the model, and a 25%testing sample for validating the model.

Firstly, withthe training data, the stepwise technique in logistic regression analysis is performed to select variables. Logistic regression is a widely used statistical model for analyzing binary outcomes, and it can makethe prediction of the odds and the related probability of an outcome or event from a set of predictor variables. In this study, the outcome is "if the family owns any guns". The predictors can be either continuous variables, categorical variables, or both.More explanation of the logistic regression model is provided below:

- The general formula of logistic regression is: ln(odds of an event occurring)= $\ln\left(\frac{p}{p-1}\right) = \beta + \beta_1 * X_1 + \beta_2 * X_2 + \cdots + \beta_n * X_n$. P is the probability of an event, which is convertible with odds.
- X_n is a predictor variable, and β_n is a regression coefficient.The relationship between the odds ratio and the coefficients is $OR = e^{\beta}$. If the coefficient β of a variable X_n is larger than 0, X_n is related to a higher odds/probability of the event. The odds ratio related to X_n is above 1 in this case.
- If the coefficient of a variable X_n is equal to 0, X_n is not related to the event.The odds ratio related to X_n is equal to 1 in this case.
- If the coefficient of a variable X_n is smaller than 0, X_n is related to a lower odds/probability of the event.The odds ratio related to X_n is below 1 in this case.

Secondly, the prediction model is tested in the testing data to examine if it provides a good prediction of the outcome. The following measures and methods are used to test the model fit:

- A receiver operating characteristic curve (ROC curve) is plotted, and the area under the ROC Curve (AUC) is reported[1]. ROC curve is a graphical plot that illustrates the diagnostic ability of a model.
- Kolmogorov-Smirnov (KS). KS statistic is a commonly used model evaluation metric for models predicting binary outcomes[2]. It tests if the logistic model separates (discriminates between) events and non-events. KS ranges from 0% to 100%, and a higher value indicates a better model fit.

Variables

Outcome variable: In the 2017 BRFSS, participants were asked "Are any firearms kept in or around your home?". 1=yes, 0=no

List of potential predictor variables is largely based on an existing publication on firearm storage[1,2]. These included demographic, socio-economic, and lifestyle factors. A total of 14 variables are entered into the logistic regression model for selection, including

- Age
- Sex
- Race/Ethnicity
- Employment Status
- Education
- Income
- MaritalStatus
- If there's any children in the family
- If the resident is a veteran
- Binge drinking
- Heavy drinking
- Smoking
- Mental health status: Number of days with no good mental health in the past month
- How often does the respondent use seat belts when driving/riding in a car. This variable may reflect a person's risk-taking behaviors or personality.

The two drinking-related variables (binge drinking and heavy drinking) were entered into the model for selection because they both are excessive alcohol intake but are different. Basically, binge drinking is drinking a lot at once, while heavy drinking is drinking a lot over a longer period.

According to the CDC[3], binge drinking is defined as when a man drinks 5 drinks of alcohol or a woman drinks 4 drinks within

1 *Evaluation of Predictive Models*. Decision Systems Group, Brigham and Women's Hospital Harvard Medical School.
2 TECHNIQUES, *M. V. MODEL VALIDATION TECHNIQUES*. Available at: https://www.listendata.com/2015/01/model-validation-in-logistic-regression.html. (Accessed: 10th February 2018)

2 hours, which can result in a blood alcohol concentration (BAC) of around 0.08. Heavy drinking is when a man has an average of 2 units of drink a day (14 a week), or 1 unit of drink per day (7 a week) for a woman.

Below is a table of the labels, names, and coding of variables.

Variable Nmber	Label	Variable Name	Coding
1	Age	X_AGEG5YR	1 Age 18 to 24 Notes: 18 <= AGE <= 24 2 Age 25 to 29 Notes: 25 <= AGE <= 29 3 Age 30 to 34 Notes: 30 <= AGE <= 34 4 Age 35 to 39 Notes: 35 <= AGE <= 39 5 Age 40 to 44 Notes: 40 <= AGE <= 44 6 Age 45 to 49 Notes: 45 <= AGE <= 49 7 Age 50 to 54 Notes: 50 <= AGE <= 54 8 Age 55 to 59 Notes: 55 <= AGE <= 59 9 Age 60 to 64 Notes: 60 <= AGE <= 64 10 Age 65 to 69 Notes: 65 <= AGE <= 69 11 Age 70 to 74 Notes: 70 <= AGE <= 74 12 Age 75 to 79 Notes: 75 <= AGE <= 79 13 Age 80 or older Notes: 80 <= AGE <= 99
2	Sex	sex	1 Male 2 Female
3	Race	X_RACE	1 White only, non-Hispanic 2 Black only, non-Hispanic 3 American Indian or Alaskan Native only 4 Asian only, non-Hispanic 5 Native Hawaiian or other Pacific Islander only, Non-Hispanic 6 Other race only, non-Hispanic 7 Multiracial, non-Hispanic 8 Hispanic 9 Don't know/Not sure/Refused
4	Employment Status	employ_status	employed unemployed homemaker/student/unable retired
5	Education	X_EDUCAG	1 Did not graduate High School 2 Graduated High School 3 Attended College or Technical School 4 Graduated from College or Technical School
6	Computed income categories	X_INCOMG	1 Less than $15,000 Notes: INCOME2 = 1 or 2 2 $15,000 to less than $25,000 Notes: INCOME2 = 3 or 4 3 $25,000 to less than $35,000 Notes: INCOME2 = 5 4 $35,000 to less than $50,000 Notes: INCOME2 = 6 5 $50,000 or more Notes: INCOME2 = 7 or 8
7	Relationship status	marital_status	married/partner never married divorced/widowed/separated
8	if there's any child in the family	with_children	1 Yes 2 No
9	If the resident is a veteran	veteran	1 Yes 2 No
10	binge drinking	binge_drinking	1 Yes 2 No
11	heavy drinking	heavy_drinking	1 Yes 2 No
12	smoking status	smoking	1 Current smoker - now smokes every day 2 Current smoker - now smokes some days 3 Former smoker 4 Never smoked
13	number of days with not good mental health in the past 30 days	mental_prob_day	1 "no bad mental health days" 2 "1-13 days" 3 ">13 days"
14	How often does the respondent use seat belts when driving/riding in a car	SEATBELT	1=always 2=nearly always 3=sometimes 4=seldom 5=never

Dataset is limited to non-missing values of all the above variables. The final dataset included 12,047 participants.

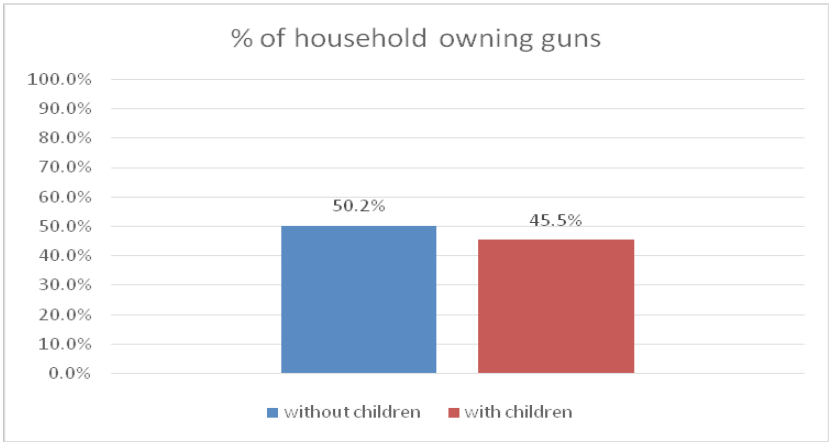
RESULTS

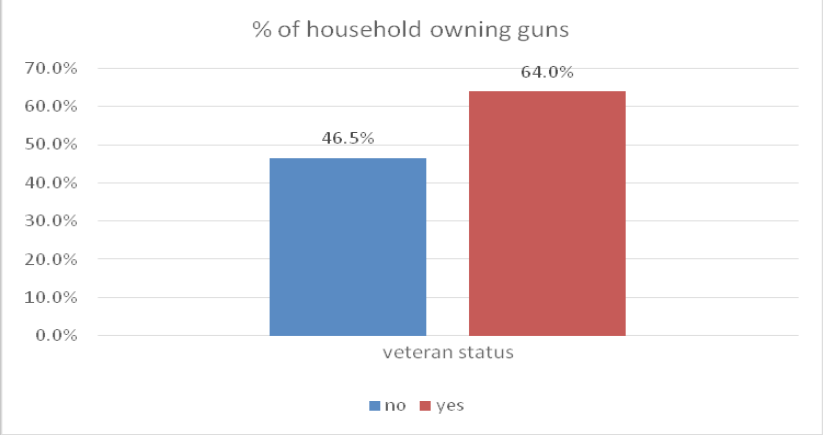
Prevalence of gun possession

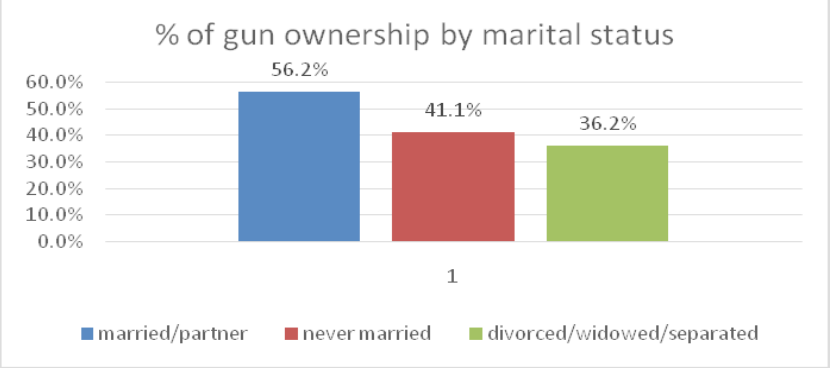
In the year 2017, there are48.8% of participants who report possessing any firearm. This proportion issimilar with national prevalence of gun ownership reported by Gallup. For example, As of 2017, Gallup found that 42 percent of American households reported possessing guns [4].

As a preliminary examination of relationship between resident/ household characteristics and gun ownership, we looked at the gun possession prevalence across race, if the family has any child, veteran status, and marital status. Other variables were not examined here, but they are also likely to be associated with gun ownership based on previous research.

data_2017_final$X_RACE	data_2017_final$firearm_ownership 0	1	Row Total
1	0.448	0.552	0.761
2	0.627	0.373	0.038
3	0.513	0.487	0.009
4	0.826	0.174	0.011
5	0.692	0.308	0.001
6	0.500	0.500	0.001
7	0.498	0.502	0.020
8	0.764	0.236	0.158





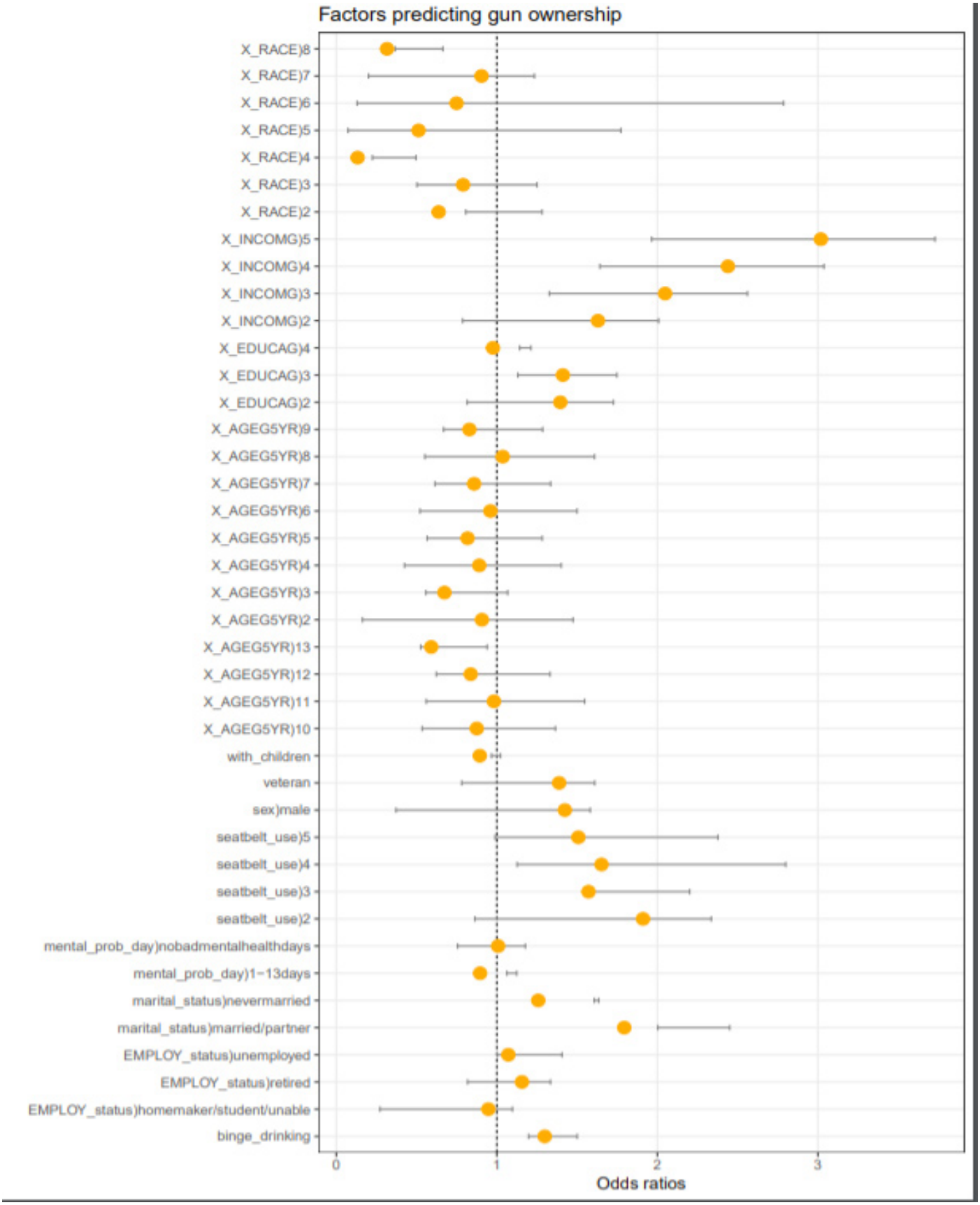


Development of the prediction model

From stepwise logistic regression analysis, 12 variables out of the 14are selected in the final prediction model. The tables of coefficients and odds ratios are listed below:

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)		Odds Ratio	lower CI	Upper CI
(Intercept)	-1.31356	0.255085	-5.149	2.61E-07	***	0.268862	0.16284	0.443
as.factor(X_AGEG5YR)2	-0.09734	0.247679	-0.393	0.694301		0.907244	0.55798	1.4748
as.factor(X_AGEG5YR)3	-0.39476	0.234714	-1.682	0.092591	.	0.673841	0.42498	1.0677
as.factor(X_AGEG5YR)4	-0.1157	0.230862	-0.501	0.616257		0.890743	0.56614	1.4012
as.factor(X_AGEG5YR)5	-0.20106	0.229424	-0.876	0.380837		0.817866	0.52125	1.2828
as.factor(X_AGEG5YR)6	-0.04055	0.227239	-0.178	0.858376		0.960262	0.61464	1.4998
as.factor(X_AGEG5YR)7	-0.15252	0.22507	-0.678	0.497989		0.858542	0.55183	1.3352
as.factor(X_AGEG5YR)8	0.036235	0.223699	0.162	0.871322		1.036899	0.66824	1.6083
as.factor(X_AGEG5YR)9	-0.18717	0.223402	-0.838	0.40213		0.829302	0.53473	1.2854
as.factor(X_AGEG5YR)10	-0.13356	0.22673	-0.589	0.555801		0.874971	0.56052	1.3651
as.factor(X_AGEG5YR)11	-0.01818	0.231622	-0.078	0.937445		0.981986	0.62313	1.5468
as.factor(X_AGEG5YR)12	-0.17743	0.236402	-0.751	0.452922		0.837418	0.52644	1.3314
as.factor(X_AGEG5YR)13	-0.52509	0.236157	-2.223	0.026183	*	0.5915	0.37198	0.9399
as.factor(sex)male	0.353406	0.053643	6.588	4.45E-11	***	1.423909	1.28186	1.5818
as.factor(X_RACE)2	-0.44993	0.120567	-3.732	0.00019	***	0.637675	0.5026	0.8065
as.factor(X_RACE)3	-0.2356	0.234643	-1.004	0.315344		0.790098	0.49684	1.2502
as.factor(X_RACE)4	-2.01875	0.284363	-7.099	1.25E-12	***	0.132821	0.0732	0.2249
as.factor(X_RACE)5	-0.66813	0.647973	-1.031	0.302487		0.512664	0.12981	1.7732
as.factor(X_RACE)6	-0.28862	0.652257	-0.442	0.658132		0.749298	0.20051	2.7859
as.factor(X_RACE)7	-0.10011	0.157865	-0.634	0.525993		0.90474	0.66413	1.2342
as.factor(X_RACE)8	-1.15178	0.076192	-15.117	< 2e-16	***	0.316073	0.27196	0.3666
as.factor(EMPLOY_status)homema	-0.05401	0.074919	-0.721	0.470977		0.947424	0.81797	1.0972
as.factor(EMPLOY_status)retired	0.145533	0.072858	1.998	0.04577	*	1.1566	1.0028	1.3343
as.factor(EMPLOY_status)unemplo	0.069247	0.138935	0.498	0.618191		1.0717	0.8153	1.4061
as.factor(X_EDUCAG)2	0.333482	0.107929	3.09	0.002003	**	1.3958	1.1307	1.7264
as.factor(X_EDUCAG)3	0.344743	0.108485	3.178	0.001484	**	1.4116	1.1422	1.7478
as.factor(X_EDUCAG)4	-0.02427	0.110225	-0.22	0.82576		0.976	0.7869	1.2124
as.factor(X_INCOMG)2	0.488825	0.105517	4.633	3.61E-06	***	1.6303	1.3274	2.0077
as.factor(X_INCOMG)3	0.717057	0.11333	6.327	2.50E-10	***	2.0483	1.6421	2.561
as.factor(X_INCOMG)4	0.891811	0.111147	8.024	1.03E-15	***	2.4395	1.9644	3.0375
as.factor(X_INCOMG)5	1.10481	0.107124	10.313	< 2e-16	***	3.0186	2.4504	3.7296
as.factor(marital_status)married/par	0.584308	0.056206	10.396	< 2e-16	***	1.7937	1.6067	2.0028
as.factor(marital_status)never marri	0.229626	0.133889	1.715	0.086337	.	1.2581	0.9668	1.6346
with_children	-0.11204	0.068668	-1.632	0.102752		0.894	0.7814	1.0228
veteran	0.328084	0.075315	4.356	1.32E-05	***	1.3883	1.1981	1.6097
binge_drinking	0.261217	0.073772	3.541	0.000399	***	1.2985	1.124	1.501
as.factor(mental_prob_day)1-13 da	-0.10975	0.087031	-1.261	0.207292		0.896	0.7555	1.0627
as.factor(mental_prob_day)no bad r	0.008295	0.079202	0.105	0.91659		1.0083	0.8633	1.1777
as.factor(seatbelt_use)2	0.64754	0.10165	6.37	1.89E-10	***	1.9108	1.5679	2.3359
as.factor(seatbelt_use)3	0.451922	0.17061	2.649	0.008076	**	1.5713	1.1266	2.2011
as.factor(seatbelt_use)4	0.502244	0.264567	1.898	0.057649	.	1.6524	0.9889	2.7995
as.factor(seatbelt_use)5	0.410704	0.231006	1.778	0.075421	.	1.5078	0.9592	2.3779



An odds ratio above 1 indicates that the variable is related to a higher risk of the event, while an odds ratio below 1 indicates that the variable is related to a lower risk of the event. According to the tables, the strongest predictor of possessing gun isthe highest income bracket (>=$50,000). The odds ratio of the predictor is 3.02, which means that the odds of a person owning a gun are increased by 3 timesif the personis in the highest income bracketcompared to a personin the lowest bracket (<$15,000).

Other predictors that make significant contributions to the modelare

- Race
- Sex
- Employment status
- Education level
- Marital status
- Veteran status
- Binge drinking
- Seatbelt use habit

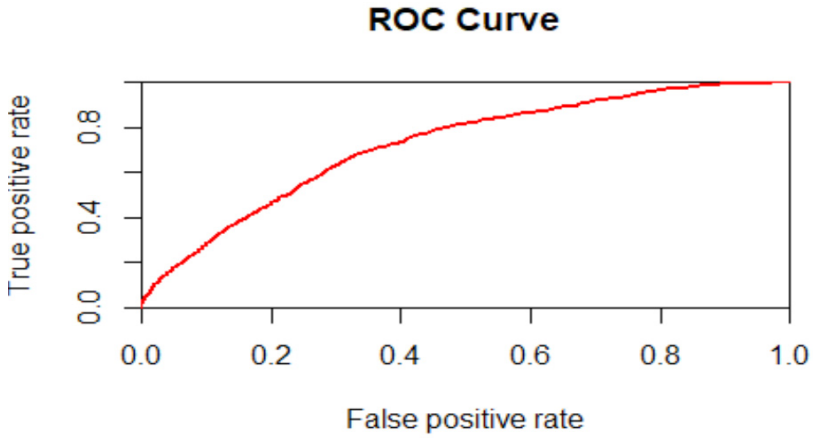
For example, the odds ratio for the predictor "veteran" is 1.38, meaning that the odds of owning gunsof a veteran is 1.38 times higher than a non-veteran.

In the training data, The AUC is 72%, which indicated that accuracy of the model is 71%.

Validation of the prediction model

The above model with 12 predictor variables is verified in the testing data. The KS statistic for the validation is 0.35. Meanwhile, the following ROC curve is generated and the AUC is 72%.

Both the KS statistic and AUC are popular metrics used to test if a model is a good fit[3]. The KS statistic measures the ability of a model to separate yes or no status of outcome events. It is suggested by researchers that KS values greater than 20% are considered acceptable for a model[5]. AUC is an estimate of the discriminatory performance of the model.In this study, a KS of 0.35 and an AUC of 72% in the validation sample indicates good performance of this model, meaning that it provides a good prediction of gun ownership.



DISCUSSION

The Washington Posthas commented that "On gun ownership, the United States stands out among the world's wealthiest nations, with an ownership rate more than three times higher than the rate in the next-highest country, Canada."[6]With the high prevalence, studies on resident and household characteristics that are related to gun ownership can be helpful in understanding what families are more likely to choose to possess firearm.

The factors identified in this study are similar with those suggested by literature. For example, using BRFSS 2004 data, Hamilton et al. discovered that "Men, veterans, middle-aged adults, non-Hispanic whites, persons with intermediate levels of education, married persons, and households without children all remain the most likely to have a gun in the home" [1]. In this study, it was also found that men, veterans, non-Hispanic whites, education levels, and marital status are associate with gun ownership. Although the variable "having children in the family" was not statistically significant in the model, from the cross tabulation we did notice that families with children had lower proportion of reporting gun possession (45.5% vs. 50.2%).

Study limitation: previous research has found that living environment is an important factor, such as urbanicity[7]

3 TECHNIQUES, *M. V. MODEL VALIDATION TECHNIQUES*. Available at: https://www.listendata.com/2015/01/model-validation-in-logistic-regression.html. (Accessed: 10th February 2018)

and neighborhood safety. The information, however, is not available in the 2017 BRFSS. A model with more comprehensive list of factors will provide even more accurate prediction of gun ownership.

For families that possess guns, it is important to keep guns in a safe manner. With this model, families that are more likely to possess guns can be identified and any safety education can be provided if necessary.

Future studies: A more comprehensive/accurate model can be achieved if living environment information is available. Meanwhile, future research can study factors associated with gun storage practices in the United States, for example, if the gun is loaded and/or locked.

CONCLUSION

A predictive model of gun ownershipin U.S. households was developed and validated. This kind of model can be helpful inidentifying residents and households that are more likely to possess firearm and to provide any education on safe gun storage practices.

REFERENCE

1. WHO (2014), *Electronic nicotine delivery systems: FCTC/COP/6/10 Rev.1*, Moscow: World Health Organization, Conference of the Parties to the WHO Framework Convention on Tobacco Control, Sixth session, 13–18 October, 2014.
2. Rahman MA, Hann N, Wilson A, Worrall-Carter L (2014). *Electronic cigarettes: patterns of use, health effects, use in smoking cessation and regulatory issues*. Tob Induc Dis. 12 (1): 21. doi:10.1186/1617-9625-12-21. PMC 4350653. PMID 25745382.
3. *DrugFacts: Cigarettes and Other Tobacco Products*. National Institute on Drug Abuse. May 2016. Retrieved 29 May 2016.
4. Deeming Tobacco Products to Be Subject to the Federal Food, Drug, and Cosmetic Act, as Amended by the Family Smoking Prevention and Tobacco Control Act; Restrictions on the Sale and Distribution of Tobacco Products and Required Warning Statements for Tobacco Products". *Federal Register*. US Food and Drug Administration. 81 (90): 28974–29106. 10 May 2016.
5. Cullen KA, Ambrose BK, Gentzke AS, Apelberg BJ, Jamal A, King BA. Notes from the Field: *Increase in use of electronic cigarettes and any tobacco product among middle and high school students — United States*, 2011–2018. MMWR Morbid Mortal Wkly Rep. 2018;67(45):1276–1277.
6. *Evaluation of Predictive Models*. Decision Systems Group, Brigham and Women's Hospital Harvard Medical School.
7. Mirbolouk, M. et al. *Prevalence and Distribution of E-Cigarette Use Among U.S. Adults: Behavioral Risk Factor Surveillance System, 2016*. Ann. Intern. Med. (2018). doi:10.7326/M17-3440
8. TECHNIQUES, *M. V. MODEL VALIDATION TECHNIQUES*. Available at: https://www.listendata.com/2015/01/model-validation-in-logistic-regression.html. (Accessed: 10th February 2018)

Citation: Jiayu Wu, "Gun Ownership in the United States: Development and Validation of a Predictive Model", American Research Journal of Humanities and Social sciences, Vol 8, no. 1, 2022, pp. 78-85.