Multiple Regression Statistical Modeling and Inference: A Study on FEMA

Abstract

With increasing frequency of natural disasters in a changing climate, the importance of federal disaster assistance is increasingly important. We collect, clean and analyze public data released from the Federal Emergency Management Agency to assess associations between local demographic characteristics and differential granting in the Individual and Households Program. Utilizing a multiple regression model with normal errors, we study associations between median household income, percentage of white households, and disaster assistance, while holding constant for multiple confounders, including disaster type, duration of the disaster, amount of damage inspected, etc. After transforming our data to better satisfy regression model assumptions, we estimate a positive adjusted relationship between median household income and average grant amount. We also estimate a negative adjusted association between the percentage of white households and the average grant amount, at high levels of median household income. We are cautious about results due to possible dependence of observations.

1. Background and Introduction

Climate change is a significant and growing problem across the world. An important consequence of our evolving climate is an increase in natural disasters and severe storms, which in 2023 alone accounted for the displacement of 3.1 million Americans (Urban Institute, 2023), and in the 20 year period from 1999-2019, accounted for the deaths of nearly 1400 individuals a year (Sharpe et al., 2021). The United States sponsors a department responsible for granting recovery aid after such events: the Federal Emergency Management Agency (FEMA). In an age where assistance after disasters is becoming increasingly relevant, we question whether the distribution of FEMA's funds are biased in any way. If so, such biases could impact long-term safety, economic stability, and recovery.

Prior research has found positive association between socioeconomic status and amount of aid received for similar damages, while higher minority populations were correlated with reduced aid allocation within the FEMA public assistance program (Domingue et al. 2019). However, no prior research has investigated procedural equity in FEMA fund dispersion for the Individual and Households Program (IHP). It is important to fill this research gap because possible bias in distribution of funds based on socioeconomic or racial status could exacerbate the affordable housing crisis currently faced by many in the United States (McClure et al., 2024).

2. Data and Exploratory Analysis

We collect data regarding FEMA's Individuals and Households Program (IHP) using the Homeowner's Housing Assistance Program Data from OpenFEMA. FEMA collects data for every disaster involving aid applications. Their website OpenFEMA has a variety of different publicly available datasets, including IHP, that catalog claims within the different FEMA programs. When a disaster occurs, individuals in the area can file claims for damages, from which FEMA assesses the site and chooses to grant an amount (none, or some portion of assessed damage). The FEMA data is aggregated by zip code; hence, we use 2023 U.S. Census Data to pair aggregate disaster claims with zip code population demographics including race, income, and education level. Our original dataset includes 27 variables, with 66,083 cases. We remove all cases missing values for any of our initial variables of interest, or with values of 0 for total inspected damage, total approved amount, or total number of houses approved (71.9% of cases). We select only data for disasters in the years 2013-2019 to keep the demographic information relevant and avoid the complications of COVID-year data. Moreover, we selected and set aside a random 10% of the data for our test set, leaving a training set with 18,581 cases.

Our primary response variable is average approved IHP amount per household, given approval for assistance. We calculate this by dividing total IHP approved amount by total number of houses approved. This is measured in dollars, and it represents our metric for aid received. Our primary predictor variables of interest are median income per household (in dollars) and percentage of population that is white. We use these as metrics for socioeconomic status and racial makeup. Notably, Average Approved Amount has a right-skewed distribution (Fig 1 to the right).

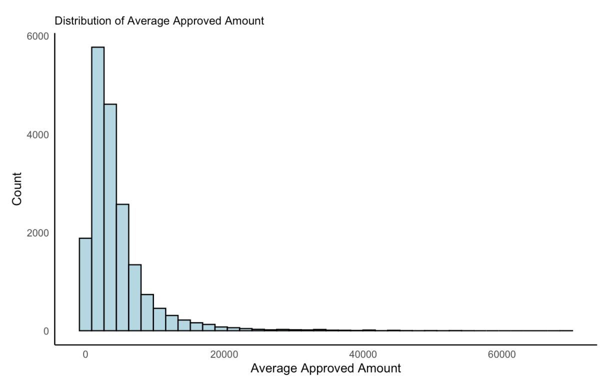


Figure 1: Histogram of continuous Average Approved Amount variable. We see an extremely right skewed distribution. The vast majority of data falls below $20,000 average approved dollars per house.

3. Model and Results

3.1 Analytical Methods

We use a normal error multiple regression model for our analysis of FEMA data. We model disaster type as a categorical variable. We use "Severe Storm" as our reference category while including other large categories (Flood, Fire, Hurricane, Severe Ice Storm, and Tornado) as indicator variables (Appendix D). We combine all other disaster types into an "Other" category represented by a separate indicator variable. We also consider an interaction term between the percent of white residents and income of the Zip Code to gain more insight into the relationship between our primary hypotheses.

When fitting a multiple linear regression of our predictors onto Average Approved Amount, we see violation of all major regression conditions. The variance of residuals increases significantly as fitted values increase. A QQ-Plot demonstrates that the residuals are not normally distributed (Appendix E). There may be some correlation in the residuals for nearby cases (i.e. zip codes in the same disaster), although a residual map of the continental United States does not show obvious violation of independence (Appendix F). Though we proceed with caution regarding independence, we conduct a $\log_{10}$ transformation of the average approved IHP amount to better satisfy the other assumptions. We also transform the total number of inspected houses to a $\log_{10}$ measure to ensure conformity to the individual linearity condition.

To further refine our model, we eliminate variables with similar motivations. We removed the percentage of black households because we already included the percent of white population in our model, and chose to only include one measure of race for simplicity. We also eliminate the median individual earnings variable because we already account for median household income (measuring a similar characteristic of the zip code). Finally, we remove the variable percent of households with children from our model, using a t-test ($p = 0.13$) to determine that it is not significant.

3.2 Final Model and Results

After transformation and variable selection, our final model assumptions are better satisfied: group linearity holds and so does linearity of the outcome variable with each individual predictor (Appendix G). There is also constant variance and expectation zero for residuals at all fitted values of average approved amount. Hence, we proceed with caution while analyzing the FEMA data in our final model. Final model variables, coefficients, 95% confidence intervals, p values, t values, and standard error are below in Figure 2.

We conclude that this regression model was effective. The F-statistic is 1,122 on 17 and 18,563 degrees of freedom. The associated p-value is less than 2.2 e-16. The regression is also fairly successful, with an adjusted R-squared of 0.506. This demonstrates that the predictors in our final model explain 50.6% of the variation in average approved amount of IHP assistance.

Variable	Estimate	Std. Error	t value	Pr(>|t|)	2.5 %	97.5 %
(Intercept)	1.370	3.529e-02	38.814	< 2e-16	1.300596	1.438942
log_{10}(Avg Inspected Dollars)	5.696e-01	4.300e-03	132.485	< 2e-16	5.612063e-01	5.780616e-01
Houses Approved for Assistance	-5.513e-05	6.435e-06	-8.568	< 2e-16	-6.774544e-05	-4.252047e-05
Duration of Disaster (days)	-1.875e-04	9.310e-05	-2.014	0.044066	-3.699633e-04	-4.982138e-06
Population Estimate	2.923e-07	1.367e-07	2.138	0.032560	2.427258e-08	5.602461e-07
% of Houses with Resident over 60	8.894e-04	2.087e-04	4.262	2.03e-05	4.804165e-04	1.298405e-03
% of Houses with Married Couple	-1.196e-03	2.627e-04	-4.553	5.33e-06	-1.711228e-03	-6.812003e-04
Median Household Income (dollars)	1.489e-06	2.883e-07	5.164	2.45e-07	9.236067e-07	2.053859e-06
% Population White	-9.748e-05	2.572e-04	-0.379	0.704694	-6.016507e-04	4.066825e-04
% Households Receiving Food Stamps	1.223e-03	3.440e-04	3.555	0.000378	5.487869e-04	1.897370e-03
% Population with Bachelor's Degree	-6.698e-04	2.356e-04	-2.843	0.004471	-1.131617e-03	-2.080503e-04
Flood Indicator	2.026e-01	1.714e-02	11.824	< 2e-16	1.690384e-01	2.362192e-01
Hurricane Indicator	1.508e-01	1.714e-02	8.802	< 2e-16	1.172585e-01	1.844401e-01
Other (not fire) Indicator	4.028e-01	2.928e-02	13.757	< 2e-16	3.453689e-01	4.601390e-01
Severe Ice Storm Indicator	1.904e-01	1.845e-02	10.319	< 2e-16	1.542037e-01	2.265256e-01
Severe Storm Indicator	1.962e-01	1.741e-02	11.270	< 2e-16	1.621060e-01	2.303630e-01
Tornado Indicator	1.935e-01	2.298e-02	8.422	< 2e-16	1.484649e-01	2.385324e-01
Median Income·Percent White	-8.744e-09	3.448e-09	-2.536	0.011222	-1.550304e-08	-1.985667e-09

Figure 2. Final Multiple Linear Regression Model coefficients

We find statistically significant adjusted coefficients for all variables that remain in our model (Appendix H). We conclude that the percentage of white residents is associated with the amount of disaster assistance granted per homeowner. We conduct a partial F-test simultaneously on the race/income interaction term and the coefficient of percentage white. The associated p value is approximately 2.4 e-8. We estimate that the expected change in log10 (average IHP grant) associated with an increase of 1% of white households in a zip code of $0 median income is 1.489e-6 log10 dollars. Incorporating the interaction term, we estimate that for each 1% increase in white households, an increase of $1 in median household income is associated with a 8.74e-9 log10 dollar decrease in IHP assistance amount. We estimate a negative association between percent white households and FEMA assistance, and our confidence grows more robust as income increases.

We conduct a similar procedure to assess association between income and amount of disaster assistance per homeowner. The partial F-test for the interaction term and income coefficient together resulted in a corresponding p-value of 9.9 e-10. We estimate a positive adjusted relationship between average IHP grant and income for all valid racial makeups. When percent of white households is 0%, we estimate that an increase of $1 in median income is associated with an increase of 1.49e-6 in the expected/average log10 approved dollar amount. With each increase of 1% white households, however, this $1 income increase in approved log dollars decreases by 8.74e-9.

To assess the prediction performance of our model, we randomly select 10% of the total dataset that we use as a testing dataset. We build a model off of the remaining 90%, and when applying our model to the testing dataset, we achieve a Root-MSE of 0.2878 and a correlation of 0.72 between actual and predicted values.

Discussion and Conclusion

In this study, we investigate the effect of demographics on the approved amount of FEMA funds in the Individual and Household Program (IHP). We find that percentage of white households and income (among other demographic factors) are significant predictors of FEMA funding distribution. We find that median household income has a significant positive association with $\log_{10}$ average approved amount while controlling for race, disaster type, damage assessed, etc., and that percentage of population white has a very small negative association with average approved amount while controlling for household income and all other variables in the model. When we investigate the interaction between percent white and median income, we find a small, negative synergistic effect of income and percent white population on average approved amount. Although the crossover of the interaction term is outside the scope of our study, its significance suggests that the relationship between these two demographic characteristics is an important consideration to employ for future study.

A few limitations to our study complicate analysis and results. For example, our final normal-QQ exemplified a distribution that slightly deviated from normal, which may impact the appropriateness of our regression model. Next, there may be independence issues since multiple zip codes likely experience the same disaster and may be demographically related. For future study, we could employ spatial regression techniques to consider these issues. Additionally, we do not consider cases with no assessed damage or approved funding. These cases may be important to consider because these cases may exhibit socioeconomic or racial bias in FEMA disaster relief. Future work should take these limitations into account and conduct more specific analysis of specific regions, disaster occurrences, and disaster types.

References

Domingue, S. J., & Emrich, C. T. (2019). Social vulnerability and procedural equity: Exploring the distribution of disaster aid across counties in the United States. The American Review of Public Administration, 49(8), 897–913. https://doi.org/10.1177/0275074019856122

Fedor, L. (2024, January 30). Americans recognize housing affordability crisis, support new policies to fix the market and build more homes. Urban Institute. https://www.americanprogress.org/article/americans-recognize-housing-affordability-crisis-support-new-policies-to-fix-the-market-and-build-more-homes

Schnall, A. H., Law, R., Heinzerling, A., Benjamin, S. M., & D'Angelo, D. V. (2021). The epidemiology and geographic patterns of natural disaster and extreme weather mortality by race and ethnicity, United States, 1999–2018. Environmental Health Perspectives, 129(10), 107001. https://pubmed.ncbi.nlm.nih.gov/34678107/

Urban Institute. (2023, August 29). More than 3 million Americans were displaced by a natural disaster in the past year. How can we prepare? https://www.urban.org/urban-wire/more-3-million-americans-were-displaced-natural-disaster-past-year-how-can-we-prepare

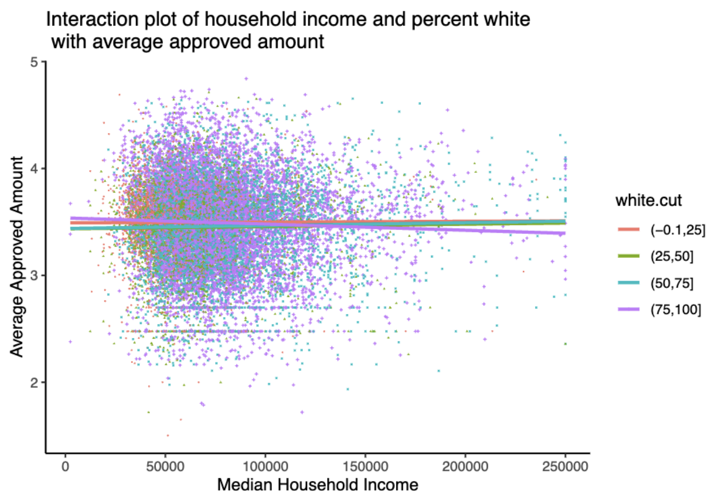
"Housing Assistance Program Data – Owners – v2." FEMA, Federal Emergency Management Agency, 15 June 2025, https://www.fema.gov/openfema-data-page/housing-assistance-program-data-owners-v2. Accessed 19 June 2025.

"S0101: Age and Sex; All 5-Digit ZIP Code Tabulation Areas within the United States." American Community Survey 5-Year Estimates Subject Tables: 2023, U.S. Census Bureau, released December 12, 2024, https://data.census.gov/table/ACSST5Y2023.S0101?q=All+5-digit+ZIP+Code+Tabulation+Areas+within+United+States+Populations+and+People. Accessed 19 June 2025.

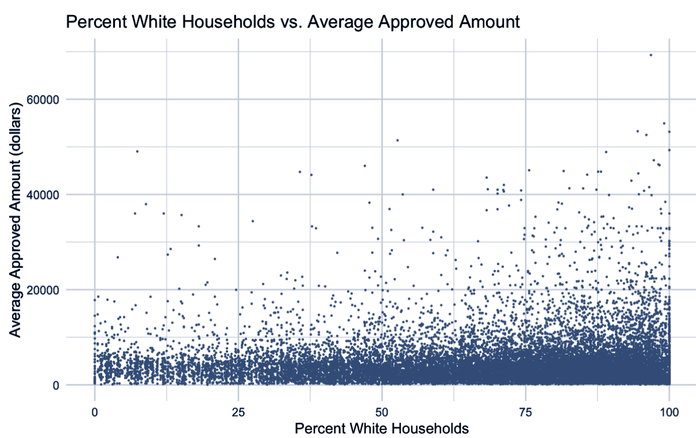
"Supplemental Nutrition Assistance Program (SNAP)." Food and Nutrition Service, U.S. Department of Agriculture, updated 13 June 2025, https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap. Accessed 19 June 2025.

Appendix

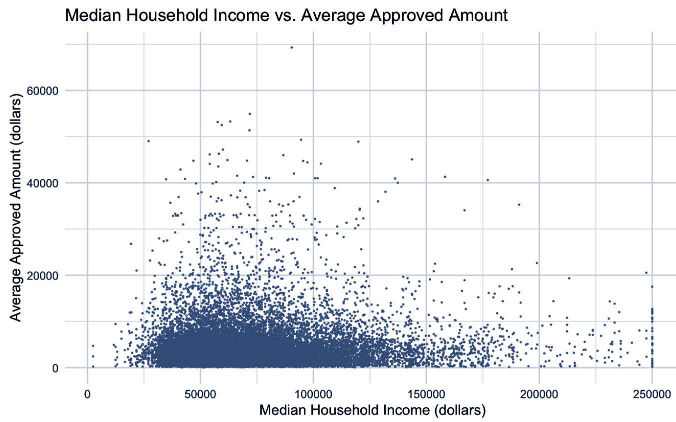
(A) Interaction plot between Median Household Income and Percentage of the Population White. We see small deviations in the relationship between income and average approved amount contingent on the percentage of the population white, with more white populations (75-100%) having a negative association, while less white populations have a positive association (0-75%).



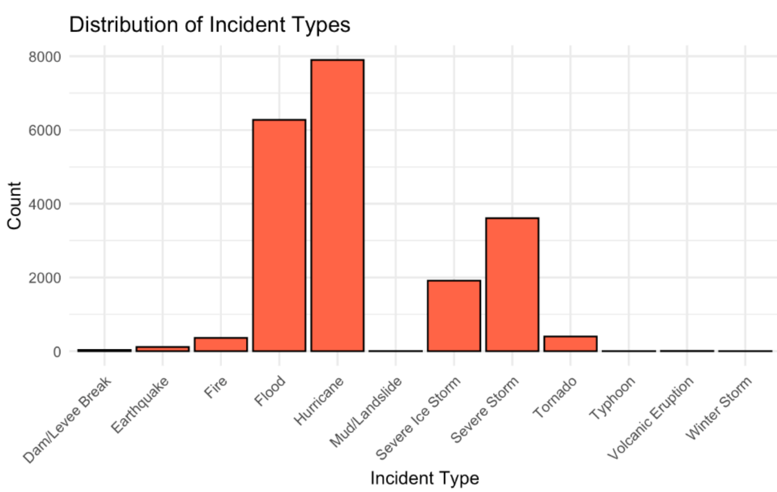
(B) Average Approved Amount vs. Percentage of White Households. We see a generally positive, though weak (r = 0.07), relationship between average approved IHP amount and the percentage of white households.



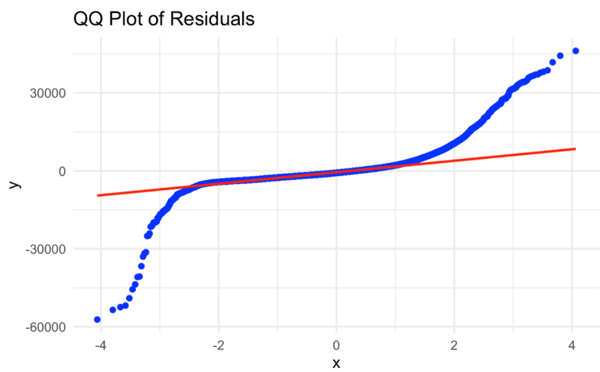
(C) Average Approved Amount vs. Median Household Income. There is a nonlinear relationship between median household income and average approved amount. Expected value and variance of approved amount increases as income increases until about $75,000, then both decrease.

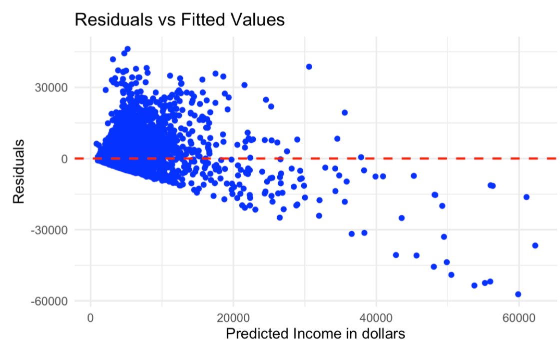


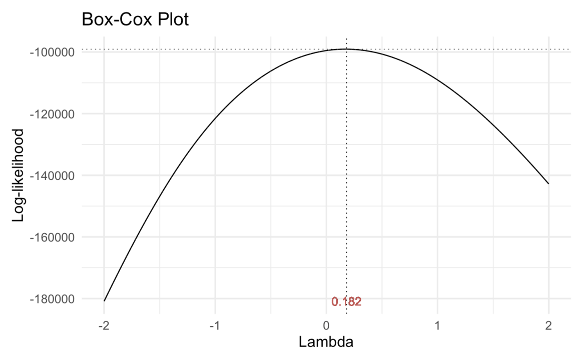
(D) Distribution of disaster types. We choose our reference category to be “Severe Storm”, because it is a large category and most “ordinary.” Indicators were created for “Flood,” “Fire,” “Hurricane,” “Severe Ice Storm,” “Tornado,” “Other”. “Other” includes Earthquakes, Volcanic Eruptions, Mudslides, Winter Storms, and Dam/Levee Break.



(E) Evaluating Regression Assumptions

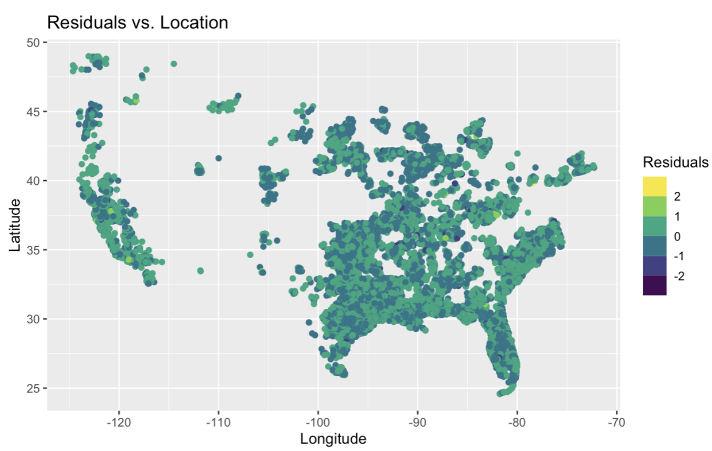


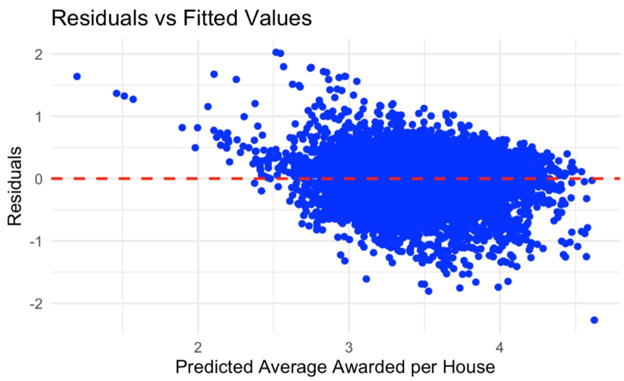


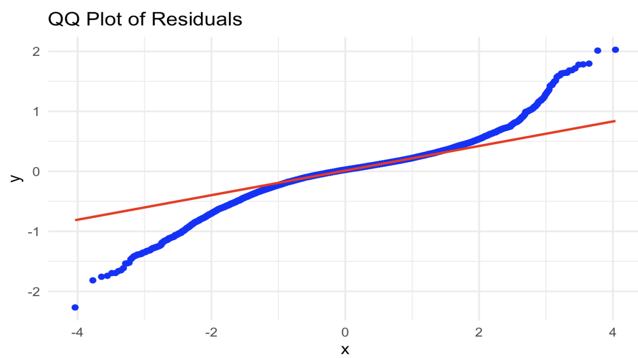


Initial diagnostic plots demonstrating violations of some regression assumptions. In the Normal-QQ Plot (top), we see severe violation of residuals' normality. In the center plot (residuals vs. fitted), we see violations of expectation 0 for residuals (expectation is less than 0 at high fitted values), and constant variance (variance increases as predicted income increases). The bottom plot, a Box-Cox plot, demonstrates that a log transformation may be beneficial. We select a log base 10 transformation on the response variable for our analysis.

(F) Residuals vs. Location plot to assess potential spatial dependence. The lack of a color pattern indicates no clear violation of the independence assumptions.







(G) We see improvements in the residual vs. fitted plot and the Normal-QQ plot above. The top graph demonstrates much closer adherence to the expectation 0 and constant variance assumptions for the residuals. There exists a small tail on the top left portion of the graph, but those points only represent about 4 of the over 20,000 remaining cases. Nevertheless, some mild violation persists in the Normal QQ plot (with the blue dots deviating from the red line at both high and low values). Although the residuals are not perfectly normal (and, not visible in the graphs, there are some independence concerns), we proceed with caution in our regression.

(H) The Final Model

$$\begin{aligned}\log_{10}(\widehat{\text{Avg Approved Dollars}}) &= 1.37 + 0.57\log_{10}(\text{Avg Inspected Dollars}) - 5.51\cdot 10^{-5}(\text{Houses Approved for Assistance}) \\ &- 1.88\cdot 10^{-4}(\text{Duration of Disaster (Days)}) + 2.92\cdot 10^{-7}(\text{Population Estimate}) \\ &+ 8.89\cdot 10^{-4}(\text{Percent of Houses with Resident over 60}) - 1.20\cdot 10^{-3}(\text{Percent of Houses with Married Couple}) \\ &+ 1.49\cdot 10^{-6}(\text{Median Household Income (Dollars)}) - 9.75\cdot 10^{-5}(\text{Percent Population White}) \\ &- 8.74\cdot 10^{-9}(\text{Median Household Income}\cdot\text{Percent Population White}) \\ &+ 1.22\cdot 10^{-3}(\text{Percent Households Receiving Food Stamps}) - 6.70\cdot 10^{-4}(\text{Percent Population with Bachelor's Degree}) \\ &+ 0.20(\text{If Incident is Flood}) + 0.15(\text{If Incident is Hurricane}) + 0.19(\text{If Incident is Severe Ice Storm}) \\ &+ 0.20(\text{If Incident is Severe Storm}) + 0.19(\text{If Incident is Tornado}) + 0.40(\text{If Incident is Other (not fire)})\end{aligned}$$