Modelling the Relationship between the Financial Damage and the Number of People Displaced by Fires

Leuven Wang

Initially Completed February 12, 2021, Revised March 4, 2021

Model

This investigation will attempt to model the relationship between the estimated number of people displaced by fire incidents and the average estimated financial damage as based on data provided by the Toronto Fire Service. We will treat the number of people displaced as the independent variable and aim to predict the corresponding average financial damage as the dependent variable. Our model will be accomplished by following the mathematical formula below:

$$Y_i = \alpha + \beta x_i + U_i$$

This particular formula generates a linear regression model presentable as a straight line of best fit. This is appropriate for our investigation as all the variables involved are numerical. We also expect there to be a clear correlation between the variables as common sense suggests that more people being displaced equates to a more widespread fire and thus, more damage and higher costs.

Here, Y_i represents the estimated average financial damage projected by the model. x_i is the number of people displaced - a variable that we know. α is the point of interception the model will have with the y-axis - that is, it is the expected average amount of damage for a fire incident that doesn't displace anyone. β is the gradient of the slope - it directs how strong and in which direction the correlation between the two variables is. U_i is an independent random variable with a normal distribution of mean 0. It is meant to account for the random fluctuations and deviations from the standard regression line of best fit that points on either side of the line will have.

Results

The variables being measured are self-explanatory. Our independent variable measures the estimated number of people displaced by the fire. Displacement is defined as the process whereby people are forced out of a property by the fire. It does not only applies to residential properties e.g. school students can be displaced if their school catches fire. Our dependent variable is the mean estimated financial damage, in Canadian dollars, of all fire incidents with that fixed number of people displaced. For the sake of simplicity, we are limiting our investigation to cases where less than 100 people were displaced, with average damage not exceeding 150,000 CAD.

Our program concludes that the following parameters are the best fitting for using linear regression to model the relationship between the number of people displaced by a fire incident and the average financial cost:

$\hat{\alpha}$	73608.8
$\hat{\beta}$	-683.9

From figure 1 it would seem that there exists a negative correlation between the number of people displaced by a fire and the average financial damage it caused. This is because β is a negative number. This means that as the number of people displaced increases, we can expect the average financial damage of the fire to drop. We can also conclude that the average cost for a fire that displaced no people is around 73,000 CAD as that is where the intercept is. If fires cost greater damage the fewer people were displaced, then the cost of damages would maximize if 0 people were displaced - since you can't displace a negative number of people. This suggests that the model would predict no higher financial damage than 73,000 CAD.
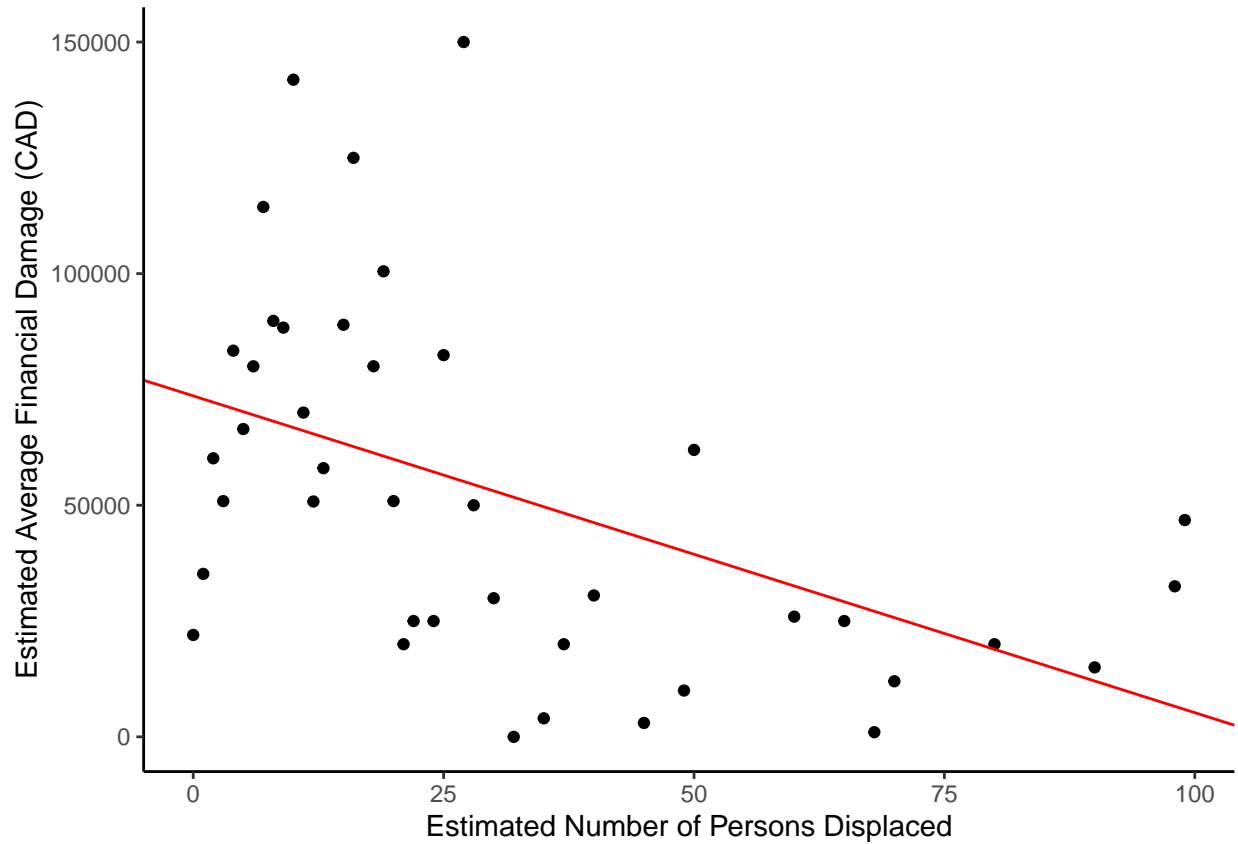

Fig 2: Linear Regression Model of Estimated Average Financial Damage Superimposed over sample data

Figure 2 visually illustrates the predictions of our model. As can be seen, the relationship between the estimated number of people displaced by a fire and the average financial damage it would cost is negative as the slope is heading downwards. This is surprising to us as we expected a positive relationship given the logical deductions we have made above about how larger displacements equate to more widespread fires and higher costs of damage. However, there are a few reasons to suggest the result we received. Firstly, notice that the sample data shows a positive correlation between the two variables in the interval where only 0 to 10 people displaced. I suggest that this is because a majority of residential fires happen within this range, thereby causing more damage to valuable possessions. Secondly, fires where more people are displaced are probably more likely to be in common areas such as auditoriums, theaters, parks etc. . . The cost would be lower as the material damaged is cheaper. These results suggest that a more complex model such as the quadratic formula would be a better method to model this particular relationship.

All analysis for this report was programmed using `R version 4.0.2`. Libraries used: openintro, tidyverse, patchwork, tidyr, dplyr

Bibliography

1. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: January 15, 2021)

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: January 15, 2021)

4. (2021, February 8) *Fire Incidents.* https://open.toronto.ca/dataset/fire-incidents/. Fire Services (Last Accessed: February 8, 2021)