An analysis of fire incidents: What are the variables that could have a relationship with the number of responding personnel

Assignment 2

Ruizhe Huang - 1006444331

25/10/2021

Introduction

The fire could be destructive sometimes. We need to study how to make reasonable decisions during fire incidents. As the residents of Toronto, we can hear the fire alarm and siren of the fire engine frequently. It would be interesting to analyze the variables that could impact the number of responding personnel. Therefore, my research question is: **What are the variables that could have a relationship with the number of responding personnel.**
I got a dataset of the fire incidents dataset from the City of Toronto Open Data Portal. It is a CSV table dataset including fire incidents as defined by the Ontario Fire Marshal. (Open Data Dataset, 2021) The sample size is subject to change yearly.
The dataset is a CSV table about fire incidents with 43 aspects imported from the City of Toronto Open Data Portal. I will use this dataset to mainly analyze the variables that could have a relationship with the number of responding personnel.
I hypothesize that the extent of fire would change the impact of fire under control time to the number of responding personnel. Since by practical rationale, we know that if the extent of fire is spread out, and with increasing fire under control time, we expected to have more responding personnel to deal with it.
In addition, I will analyses the numerical and graphical summaries for the extent of fire, under control time and number of responding personnel. The extent of the fire is essential to determine whether the fire is severe. If the fire is not confined to the object of origin, it could be more serious. The under control time is essential since it measures the time the fire stations get notified until they control the fire. With longer under control time, we may have more property loss or casualties. Lastly, the number of responding personnel gives us an idea of how many people will fight the fire. Knowing the results could help the fire commander or the people at the fire scene decide whether the extent of the fire is confined to the object of origin. The fire commander or the people at the fire scene could rapidly survey the scope of the fire, working out an effective plan to control the fire. Therefore, we have introduced the importance of this analysis.
In the data section, there will be a data description in detail. Moreover, I have numerical summaries and graphical summaries of the variable of interest: the extent of fire, under control time and number of responding personnel.
In the methods section, I will introduce the multiple linear regression model explained in detail.
In the results section, I will find the estimated multiple linear regression model of extent of fire, under control time and number of responding personnel.
In the end, there will be a conclusion to summarize the main points of this report.

Terminology

TFS: Toronto Fire Services is the City's only all hazards emergency response organization. Fire Services provides Toronto residents, visitors and businesses with protection against loss of life, property and the

environment from the effects of fire, illness, accidents, and all other hazards through preparedness, prevention, public education, and emergency response, with an emphasis on quality services, efficiency, effectiveness, and safety. (City of Toronto, 2020)
CLT: The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution. (Ganti, 2021)
The other important terminologies will be described in the data variable description part.

Data

Data Collection Process

I collected the fire incidents dataset from the City of Toronto Open Data Portal. The website URL of the source is https://open.toronto.ca/dataset/fire-incidents/. It is a dataset including fire incidents as defined by the Ontario Fire Marshal. The type of this dataset is a CSV table. This dataset was refreshed annually, and the last date refreshed was on Oct 18, 2021. (Open Data Dataset, 2021)

Data Summary

The dataset is from the City of Toronto Open Data Portal. It is a CSV table dataset including fire incidents as defined by the Ontario Fire Marshal. (Open Data Dataset, 2021) The original dataset contains 17536 observations and 43 variables. Each observation, in this case, represents a fire incident. That is, it contains 17536 fire incidents along with various other statistics for each fire incident such as extent of fire, number of responding personnel, number of responding apparatus and Sprinkler System Operation and so on, in total 43 different aspects, one of them is the unique id for each fire incident.

Data Cleaning Process

Since the dataset was imported from the website and it was unorganized with lots of missing values and some variables that are of no use to my analysis, I will perform data cleaning at first.
I remove the missing values for the following variables *Extent Of Fire*, *Fire Under Control Time*, *TFS Alarm Time*, *Number of responding personnel* and *under control time*. After that, I have checked that the numerical summaries are not changed a lot.
A new variable *under control time* was created. It represents the time difference, in minutes, of the time of fire under control and the time Toronto Fire Service was notified of the incident. It shows the time it takes to control the fire more directly.
Moreover, since the variable *Extent Of Fire* has multiple level values. For instance, Confined to object of origin, Confined to roof/exterior structure, Confined to part of room/area of origin, Spread beyond room of origin., etc. I changed it into the binary response: Confined to object of origin or Not confined to object of origin.
Then, variables that were not useful was removed from the dataset, and only 8 critical variables were kept, including *id*, *Extent Of Fire*, *Number of responding apparatus*, *Number of responding personnel*, *Fire Alarm System Operation*, *Smoke Alarm at Fire Origin*, *Sprinkler System Operation* and *under control time*. Since these variables are the main factors for a fire incident, I will study the relationship between the under control time, extent of fire and number of responding personnel later in this report.
The cleaned dataset contains 11210 observations and 8 variables without any missing values.

Variable Description

Here are the important columns within the cleaned dataset.

Table 1: Description of important variables

Variable name	Description	Type
Extent Of Fire	whether extent of fire is confined to the object of origin	Categorical
Fire Alarm System Operation	The status of fire alarm system operation	Categorical
Number of responding personnel	Number of responding personnel in each fire incident	Numerical
Number of responding apparatus	Number of responding apparatus in each fire incident	Numerical
under control time	The time difference of the time of fire under control and the time TFS was notified of the incident in minutes	Numerical

Table 1 is a description of selected essential variables. I kept only a few variables to give readers a picture of the critical factors for each fire incident.

Extent Of Fire represents whether extent of fire is confined to the object of origin. It contains 'Confined to object of origin' or 'Not confined to object of origin' as the response.

Fire Alarm System Operation representing the status of fire alarm system operation. It has multiple levels. For instance, Fire alarm system operated, Not applicable (no system), Fire alarm system operation undetermined, and Fire alarm system did not operate.

Number of responding personnel representing the number of responding personnel, including First Class Firefighter, Captain, Division Commander., etc. Those who have to receive intermediate and advanced level chemical, biological, radiological, nuclear and explosives(CBRNE) training administered by the Royal Canadian Mounted Police(RCMP) under Public Safety Canada. (Toronto fire services master fire plan, 2019)

Number of responding apparatus: representing the number of responding apparatus that is Frontline Apparatus (Staffed) such as Pumper, Aerial, Ladder, Tower and High Rise Unit., etc.

Chief units and Command Vehicles such as Fire Chief, Division Commander, Platoon Chief and District Chief., etc.

Support Apparatus (Cross-staffed) such as Hazmat Support Unit, Decontamination Unit, Water Tanker, Rapid Attack Vehicle, Fireboat, Mechanical Response Unit and All-Terrain Vehicle., etc.

Miscellaneous Apparatus such as Training Pumper, Spare vehicles and Fire Investigator., etc. (Friebe & Lasiuk, 2003)

under control time representing the time difference of the timestamp of fire under control and the timestamp Toronto Fire Service was notified of the incident in minutes. It shows the time it takes to control the fire since the Toronto Fire Service was notifies.

This report will only focus on the following variables: *Extent Of Fire*, *Number of responding personnel* and *under control time*. These variables are essential since I intended to use these variables to analyze the relationship between the under control time, extent of fire and number of responding personnel.

Numerical summaries

Table 2: Numerical summaries of two numerical variables

Variables	Mean	Median	Standard Deviation	Min	Max	IQR
Number of responding personnel	30.21	22	26.12	2	1275	15
under control time	19.76	11	49.41	1	2015	12

Table 2 is the numerical summaries of the number of responding personnel and the fire under control time(in minutes). There will be some graphical summaries analyzing these variables later as well.
From Table 2, we know that among 11210 fire incidents, the sample mean and sample median of the number of responding personnel is 30.21 and 22, respectively. So, It is right-skewed. The sample standard deviation of the number of responding personnel, which is 26.12, it expected a large fluctuation of the data. The sample range of the number of responding personnel is from 2 to 1275, which is a really extensive range. The sample IQR of the number of responding personnel is 15, which means that the range between the 1st and the 3rd quartile is 15, which is a relatively large spread.
For the fire under control time, the sample mean is 19.76 minutes, which is higher than its sample median of 11 minutes. It implies that the fire under control time is right-skewed. The sample standard deviation of the fire under control time is very significant at 49.41 minutes. It would expect a relatively large fluctuation of the data. The sample range is between 1 and 2015 minutes, which is a fairly large range. The sample IQR is 12 minutes, which means that the range between the 1st and the 3rd quartile is 12 minutes, which is a relatively small spread.
In summary, the fire under control time has a relatively more considerable fluctuation and spread. That means practically the time fire is controlled for each fire incident varies a lot.

Table 3 is a proportion table of extent of fire (From 11210 observations).

Table 3: Proportion of extent of fire

term	Confined to object of origin	Not confined to object of origin
proportion	45.69%	54.31%
count	5122	6088

Table 3 is a numerical summary of the proportion of different extent of fire proportion. From 11210 observations in the cleaned dataset, we have 5122 fire incidents confined to object of origin and 6088 fire incidents is not confined to object of origin. It is clear to see that fire not confined to object of origin account for the most significant proportion(54.31%), which is around 1.2 times of the fire confined to object of origin(45.69%). There is a graphical analysis in the next part to visualize the proportion of the extent of fire as well.

Graphical summaries

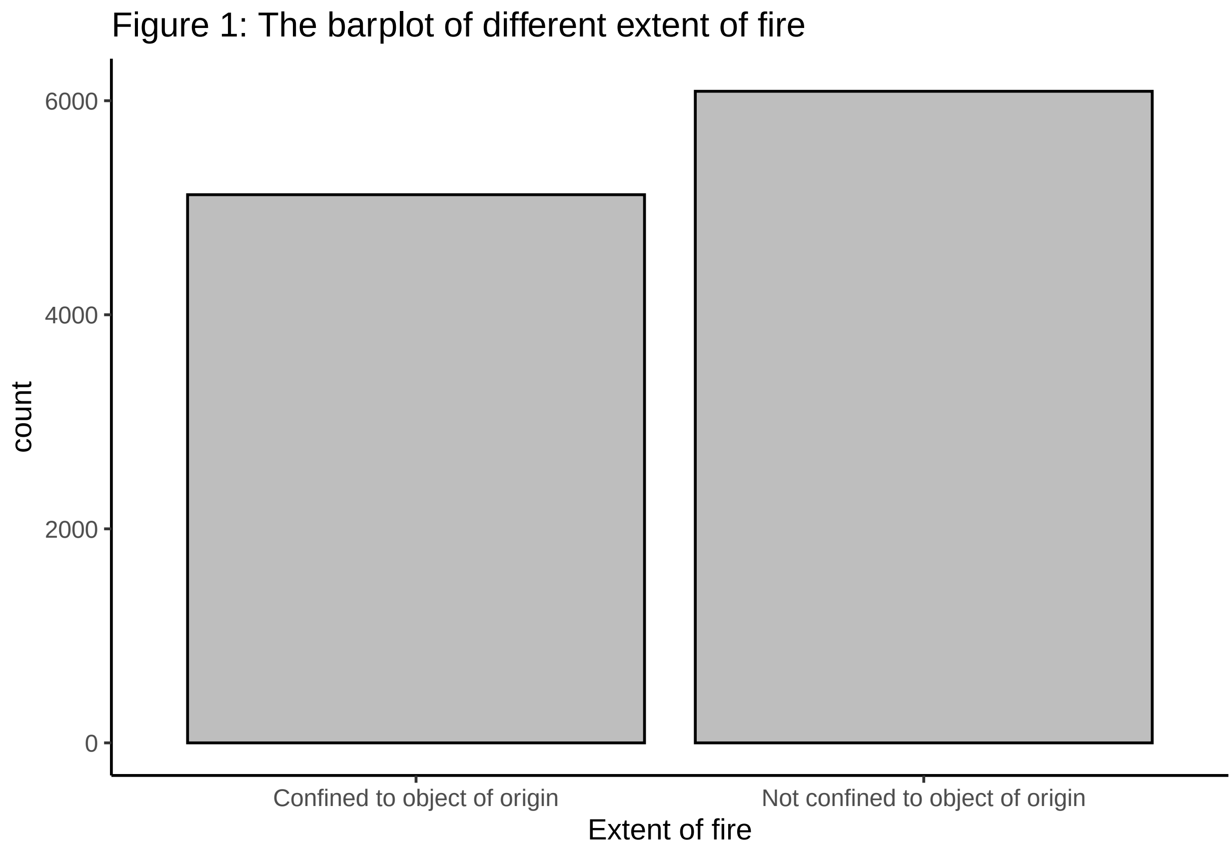

Figure 1 is a barplot of the extent of fire(categorical variable). The fire not confined to the object of origin accounted for the most significant proportion. The fire confined to the object of origin accounted for a relatively small proportion. The graph shows that nearly 6000 fire incidents are not confined to objects of origin, and around 5000 fire incidents are confined to objects of origin. The number of fire incidents not confined to the object of origin is about 1.2 times that of fire confined to the object of origin. Hence by figure 1, we found that fire not confined to the object of origin occur more frequently than fire confined to the object of origin.

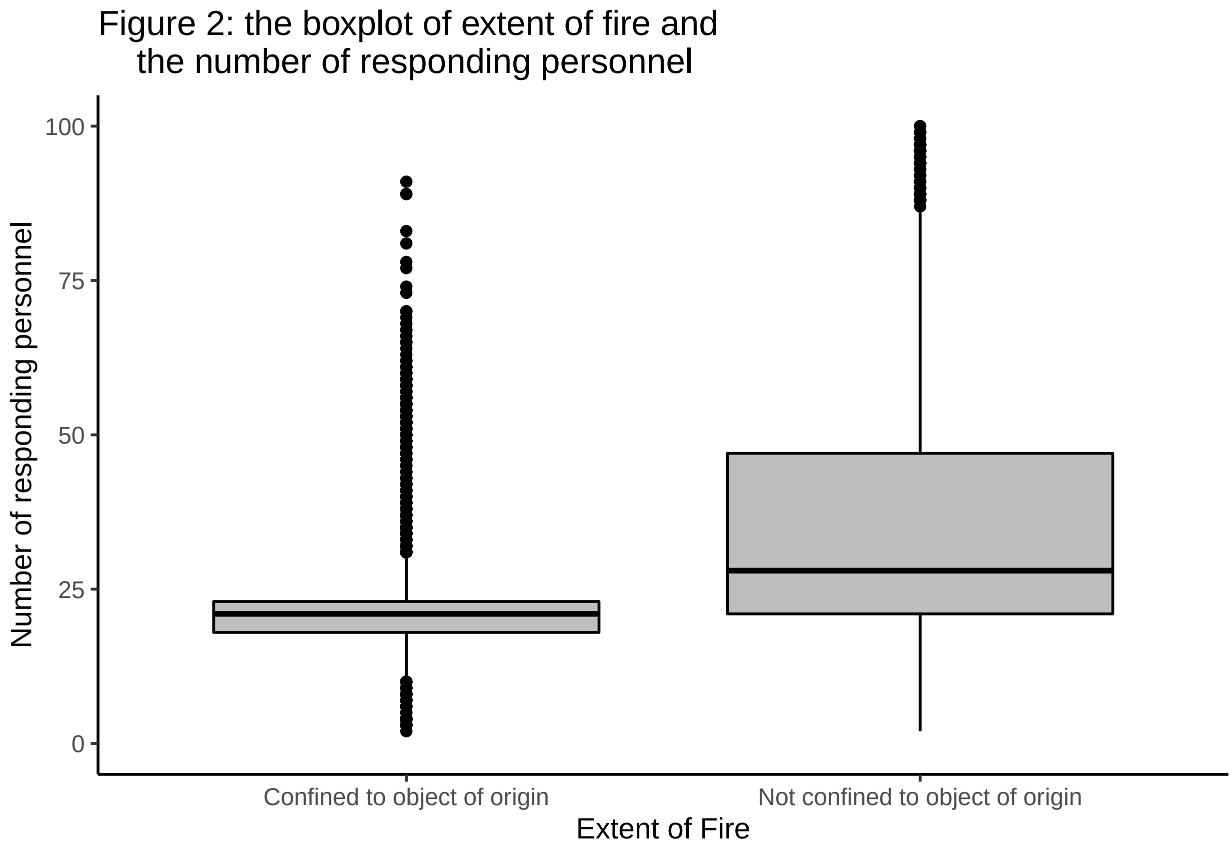

Figure 2 shows the boxplots of extent of the fire and the number of responding personnel. Since the range of this variable is from 2 to 1275, and there are a few extreme values in this variable. However, these extreme values indeed exist in real life. We cannot just remove it from the dataset. It is hard to visualize the boxplot with these extreme values. Therefore, for graphing, I only show the numbers of responding personnel smaller than 100 as the x-axis, but those extreme outliers are still in the dataset.

For the fire confined to the object of origin, the boxplot is almost symmetric with lots of outliers under 10 responding personnel and above 30 responding personnel. The median number of responding personnel when the fire is confined to an object of origin is centred around 20. The first quantile nearly at 17. And the third quantile nearly at 23. So the IQR of the number of responding personnel when the fire is confined to the object of origin is approximately 6. It implies that the spread of the number of responding personnel when the fire is confined to the object of origin is relatively small.

The boxplot is right-skewed with lots of outliers above 80 responding personnel for the fire that is not confined to the object of origin. The median of the number of responding personnel when the fire is not confined to its origin is centred around 27. The first quantile is nearly at 20. And the third quantile nearly at 40. So the IQR of the number of responding personnel is approximately 20. It implies that the spread of the number of responding personnel when the fire is not confined to the object of origin is relatively large.

In conclusion, the spread of the number of responding personnel when the fire is not confined to the object of origin is much larger than fire confined to the object of origin. However, the median number of responding personnel for both confined or not confined to the object of origin is almost the same. Therefore, we would expect that we would have nearly the same number of responding personnel regardless of whether the fire is or is not confined to the object of origin most of the time. However, for the fire not confined to the object of origin, we would expect significant fluctuation for the number of responding personnel. Most likely, we would expect more responding personnel.

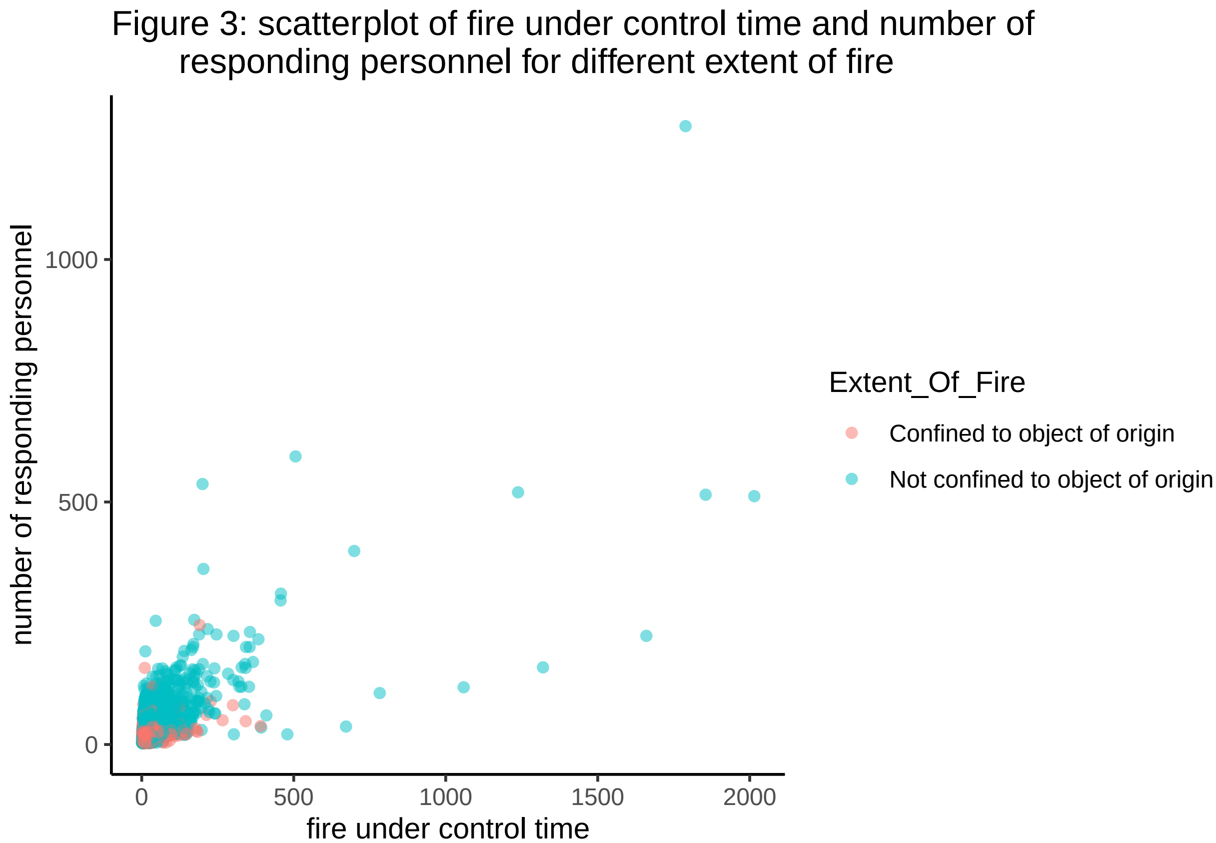

Figure 3 is a scatterplot that demonstrates the relationship between the fire under control time(numerical), extent of fire(categorical) and number of responding personnel(numerical). We can see a different trend for the extent of fire. For the fire not confined to object of origin, the number of responding personnel is greater for the same under control time than confined to object of origin. It means that when the fire spreads out to its origin, we need more responding personnel to control the fire.

Moreover, the points are concentrated tightly for the fire under control time lower than 500 minutes. It may form a strong positive linear relationship between fire under control time, extent of fire and the number of responding personnel. Based on these observations, the multiple linear regression model seems appropriate for these variables. We will see how exactly the multiple linear regression model is like in the method and result section.

Wrapping up for data section

I want to emphasize several critical results in the data section. Table 2 found that the fire under control time has a relatively more considerable fluctuation and spread. According to table 3 and figure 1, we have the fire that is not confined to the object of origin account for the most significant proportion(54.31%), which is around 1.2 times of the fire confined to the object of origin(45.69%). Hence, fire not confined to the object of origin occur more frequently than fire confined to the object of origin. Furthermore, from figure 2, we would expect that we would have nearly the same number of responding personnel regardless of whether the fire is or is not confined to the object of origin. However, for the fire not confined to the object of origin, we would expect significant fluctuation for the number of responding personnel. Most likely, we would expect more responding personnel. From figure 3, we can see a different trend for the extent of the fire. For the fire not confined to object of origin, the number of responding personnel is greater for the same under control time than confined to object of origin. It may form a strong positive linear relationship between fire under control time, extent of fire and the number of responding personnel. We will see how exactly the multiple linear regression model is like in the method and result section.

Methods

To study the relationship between the fire under control time(numerical), extent of fire(categorical) and number of responding personnel(numerical) as hypothesized in the introduction part. We have shown in the data section that the multiple linear regression model seems appropriate for these variables. I will use the Frequentist linear regression to analysis it. Assume regression coefficients are constant, and response variable is a numerical variable.
Here is the mathematical model for the multiple linear regression:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,A} + \beta_3 X_{i,1} \times X_{i,A} + \epsilon_i$$

$i = 1, \ldots, n$ where n is the number of fire incidents in the sample.
Y_i is the number of responding personnel for the i^{th} fire incident.
$X_{i,1}$ is the fire under control time for the i^{th} fire incident.
$X_{i,A}$ is the extent of fire for the i^{th} fire incident. It equals 1 if the extent of fire for i^{th} fire incident is not confined to its origin. It equals 0 if the i^{th} fire incident is confined to its origin.
β_0 is the intercept of fire confined to its origin. It represents the expected value of the number of responding personnel for the fire is confined to its origin when fire under control time is 0.
$\beta_0 + \beta_2$ is the intercept of fire that is not confined to its origin. It represents the expected value of the number of responding personnel for the fire is not confined to its origin when fire under control time is 0.
β_1 represents when the fire under control time increase by one unit, the expected value of the number of responding personnel changes by β_1 for the fire that is confined to the object of origin.
β_2 represents the difference in the expected value of the number of responding personnel between two groups when fire under control time is 0.
$\beta_1 + \beta_3$ represents when the fire under control time increase by one unit, the expected value of the number of responding personnel changes by $\beta_1 + \beta_3$ for the fire that is not confined to the object of origin.
The mathematical model implicitly assume that a linear relationship exists in the population. Assume ϵ_i are independent and identically distributed error terms. That is $\epsilon_i \overset{iid}{\sim} Normal(0, \sigma^2)$, which implies Y_i follows Normal distribution. Another assumption is that the error are uncorrelated, namely $Cov(\epsilon_i, \epsilon_j) = 0$ or $Cov(y_i, y_j) = 0$. Last assumption is that the errors ϵ_i have a common variance σ^2. That is constant error variance, namely, homoskedasticity. (Daignault, 2021)

Model selection process

For the model selection process, since the research question is: what are the variables that could have a relationship with the number of responding personnel. Therefore, I will use the *number of responding personnel* as the response variables. Then, I will use the *Extent Of Fire* and *under control time* as the predictors hypothesized in the introduction by the practical rationale. Besides, as shown in figure 3, we can see a different trend for the extent of fire. That is, the extent of fire would change the impact of fire under control time to number of responding personnel. Therefore, I decided to include the interaction term of under control time and extent of fire. Then, check the p-value of the coefficients for the multiple linear regression model by using the p-value significance cut-off, removing the variables that are not significant, which means the p-value is larger than 0.05.
Moreover, I will show a histogram of the *number of responding personnel*(response variable) to check the assumption of Normality. Also, I will provide a scatterplot of under control time and number of responding personnel for the different extent of fire with regression lines to check other assumptions and visualize the regression lines.

Results

I selected the variables in table 4 to construct the multiple linear regression model. *Number of responding personnel* is the response variable. *Extent Of Fire* and *under control time* are the predictors. I removed the missing values in each selected variable as well.

Table 4: Description of multiple linear regression variables

Variable name	Description	Type
Extent Of Fire	whether extent of fire is confined to the object of origin	Categorical
Number of responding personnel	Number of responding personnel in each fire incident	Numerical
under control time	The time difference of the time of fire under control and the time TFS was notified of the incident in minutes	Numerical

Table 4 is the variables description table for the variables used in the multiple linear regression model. A more detailed description can be found in the data variable description section.

Table 5: Description of multiple linear regression variables

term	estimate	standard error	statistic	p-value
Intercept	19.8063779	0.3410004	58.0831584	0
under control time	0.1899735	0.0156906	12.1074435	1.561476×10^{-33}
Extent Of Fire Not confined to object of origin	8.5079156	0.430106	19.7809719	$1.2443462 \times 10^{-85}$
under control time:Extent Of Fire Not confined to object of origin	0.149274	0.0161387	9.2494391	$2.6648909 \times 10^{-20}$

Table 5 is a summary table for the coefficient of the multiple linear model. From table 5, eventually the estimated multiple linear regression model is:

$$\hat{Y}_i = 19.81 + 0.19X_{i,1} + 8.51X_{i,A} + 0.15X_{i,1} \times X_{i,A}$$

where $\hat{Y}_i$ is the expected number of responding personnel, $X_{i,1}$ is the under control time and $X_{i,A}$ is the extent of fire.

When fire under control time increase by 1 minute, the average number of responding personnel will increase by 0.1899735 minutes for fire confined to object of origin.

When fire under control time increase by 1 minute, the average number of responding personnel will increase by 0.3392475 minutes for fire not confined to object of origin.

As $\hat{\beta}_0$=19.8063779. It means when the fire is confined to its origin when fire under control time is 0, the expected number of responding personnel is 19.8063779. However, when the fire under control time is 0, there would be in average 19.8063779 responding personnel for the fire confined to its origin. It seems does not make sense in real life.

For $\hat{\beta}_0 + \hat{\beta}_2$=28.3142936. It represents the expected value of the number of responding personnel for the fire is not confined to its origin when fire under control time is 0. However, for the fire not confined to its origin, when the fire under control time is 0, there would be in average 28.3142936 responding personnel. It seems does not make sense in real life as well.

Also $\hat{\beta}_2$=8.5079156. It means the difference in the expected value of the number of responding personnel between two groups is 8.5079156, when fire under control time is 0. However, it does not make sense as fire under control time normally will not be 0.

The p-value for $H_0 : \beta_3 = 0$, $H_a : \beta_3 \neq 0$ is $2.6648909 \times 10^{-20}$. The common significant level is 0.05. It

means that when the p-value is lower than 0.05, we have strong evidence against that $H_0 : \beta_3 = 0$. That is, we have strong evidence to support that the extent of fire would change the impact of fire under control time to number of responding personnel.
Altogether, we know that the estimated model when fire is not confined to the object of origin would be:

$$\hat{Y}_i = 28.31 + 0.34X_{i,1}$$

where $\hat{Y}_i$ is the expected number of responding personnel, $X_{i,1}$ is the under control time.
Then, we know that the estimated model when fire is confined to the object of origin would be:

$$\hat{Y}_i = 19.81 + 0.19X_{i,1}$$

where $\hat{Y}_i$ is expected number of responding personnel, $X_{i,1}$ is the under control time.
Practically, when we find out that the fire is not confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.3392475. When we find out that the fire is confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.1899735.

Figure 4: the histogram of number of responding personnel

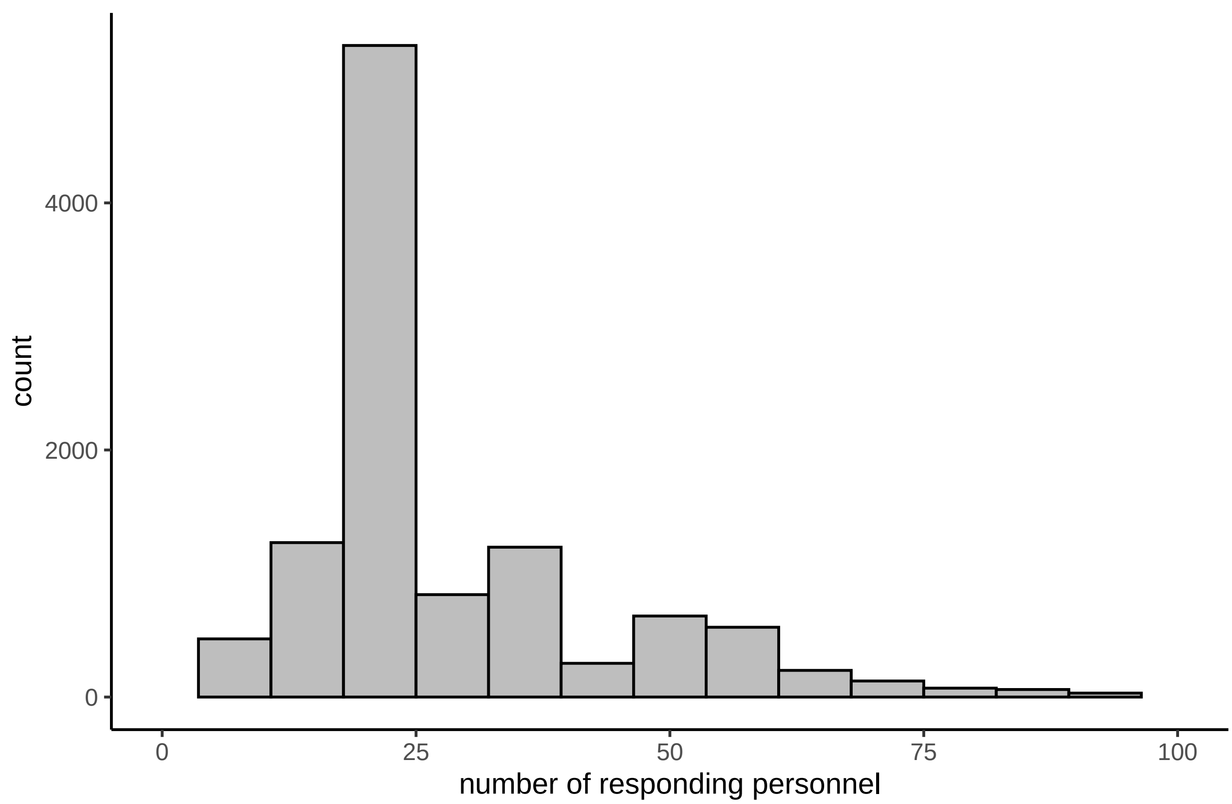

Figure 4 is a histogram of the numbers of responding personnel. Since the range of this variable is from 2 to 1275, and there are a few extreme values in this variable. However, these extreme values indeed exist in real life. We cannot just remove it from the dataset. It is hard to visualize the histogram with these extreme values. Therefore, for graphing, I only show the numbers of responding personnel smaller than 100 as the x-axis, but those extreme outliers are still in the dataset.
The shape of this histogram is right-skewed with one significant mode, which is around 25. Moreover, most of the time, we found that the number of responding personnel is between 15 and 30. The centre of the histogram is around 23. It means that we expect to have approximately 23 responding personnel present most of the time during a fire incident.
Since the number of responding personnel is used as the response variable for the multiple linear regression

model, thus, it is supposed to follow a Normal distribution, which is the assumption of the linear model. However, from figure 4, we can see that the number of responding personnel is right-skewed. So, we can apply Central Limit Theorem(CLT) here, as we have a considerable amount of data, which is more than 10000 observations. Since data points are independently and identically distributed, as the sample size(n) is large enough, the sample mean's distribution would converge to a Normal distribution by the CLT.

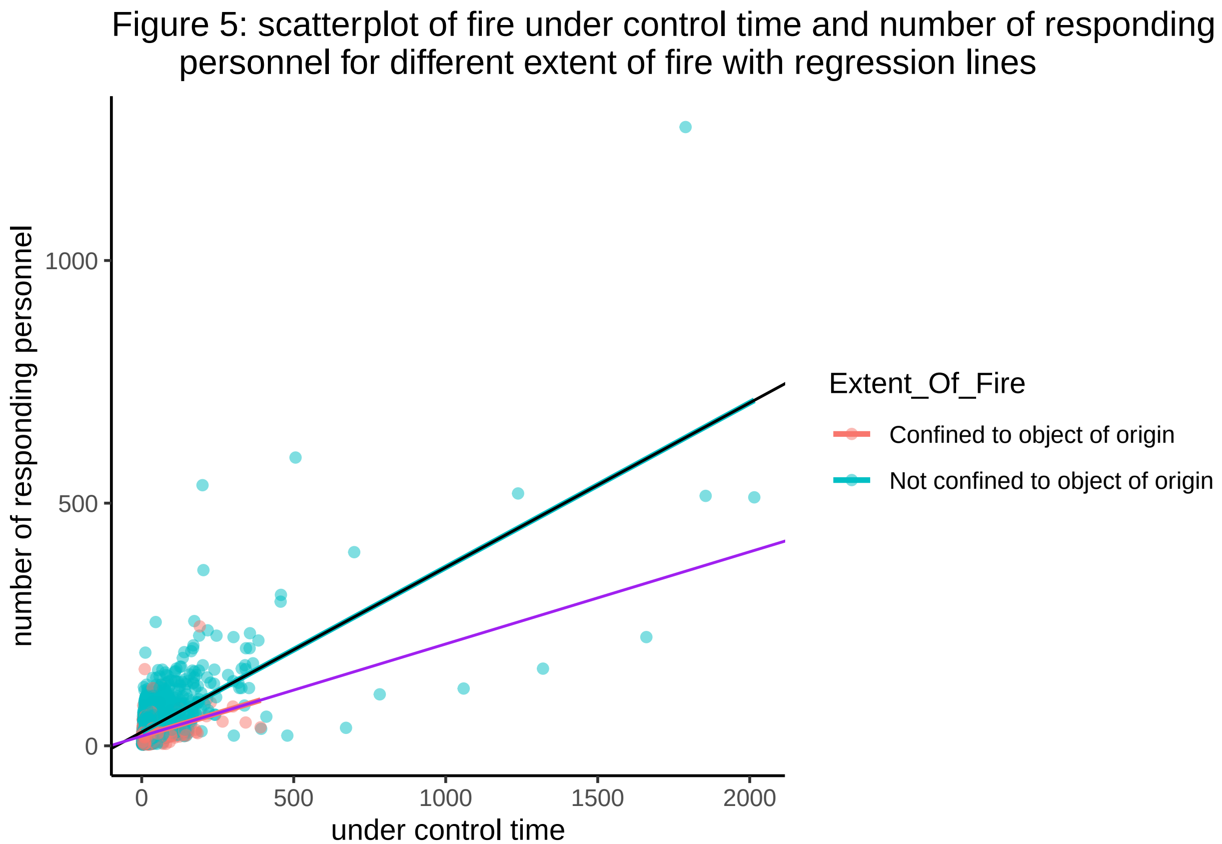

Figure 5 is a scatterplot of the relationship between the fire under control time(numerical), extent of fire(categorical) and number of responding personnel(numerical). The black line and purple line in the scatterplot visualize the linear regression line for the fire not confined to object of origin and confined to object of origin, respectively. There is a slightly strong positive linear relationship between the fire under control time and number of responding personnel for both confined to object of origin and not confined to object of origin. As fire under control time increase by 1 minute, the expected number of responding personnel will increase by 0.1899735 for fire confined to object of origin, and 0.3392475 for fire not confined to object of origin. Moreover, we can see the different slopes and intercept in figure 5. Specifically, the slope of fire confined to object of origin is lower than the slope of fire not confined to object of origin, as shown in the summary table of the multiple linear regression model. Moreover, we have discussed that the intercept of this model does not make sense in real life.

However, the relationship merely is appropriate for the fire under control time between 1 and 2015 minutes, which is the minimum and maximum fire under control time from the dataset. Since I do not have the data that when the fire under control time is lower than 1 minute or exceeds 2015 minutes, I will not know how the number of responding personnel would be like. Therefore, we cannot estimate the relationship from this plot when the fire under control time lower than minimum or exceeds the maximum value from the dataset. The linearity assumption for the linear regression model seems satisfied from figure 5, and the normality assumption has shown in figure 4. One of the linear regression model assumptions is that the random fluctuation will be around the regression line, which is validated since the points are almost evenly distributed around the regression line. Another assumption of the linear regression model assumes the same variance. Each random fluctuation is supposed to have the same amount of variability. Figure 5 somewhat violates the constant variance assumption. Since each random fluctuation is supposed to have the same variability, this is violated because the distance between the points and the regression line is not constant in figure 5. Therefore, precision in our estimate would be less, and the hypothesis test for the coefficients in the linear regression model may be less reliable.

Wrapping up for result section

In the result section, I used the practical rationale and p-value significance cut-off to determine the estimated multiple linear regression model for this report. That is $\hat{Y}_i = 19.81 + 0.19X_{i,1} + 8.51X_{i,A} + 0.15X_{i,1} \times X_{i,A}$, where $\hat{Y}_i$ is the expected number of responding personnel, $X_{i,1}$ is the under control time and $X_{i,A}$ is the extent of fire. Practically, when we find out that the fire is not confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.3392475. When we find out that the fire is confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.1899735. Moreover, I have checked the assumptions of the linear model. The constant variance assumption may be violated in the model. Therefore, precision in our estimate would be less, and the hypothesis test for the coefficients in the linear regression model may be less reliable. In addition, From figure 4, we can see that we expect to have approximately 23 responding personnel present most of the time during a fire incident. Besides, we found that the number of responding personnel is right-skewed. However, we can apply CLT as the sample size is large enough. The distribution of the sample mean would converge to a Normal distribution.
Practically, we could apply these results to the fire scene. These results could help commanders decide the plan depending on whether the fire is confined to the object of origin. If the fire is not confined to the object of origin, we would expect more responding personnel needed for every minute until the fire is controlled. Thus, I would suggest that once the commanders find out that the fire is not confined to the origin. We should rapidly survey the scope of the fire, working out an effective plan to control the fire. For the witness or the people present at the fire scene, we should try our best to prevent contact between the object of origin and combustible material, trying to confine the fire at the origin rather than spread out elsewhere.

Conclusions

As specified in the introduction, my research question is: what are the variables that could have a relationship with the number of responding personnel.
I hypothesize that the extent of fire would change the impact of fire under control time to number of responding personnel. Since by practical rationale, we know that if the extent of the fire is spread out, and with increasing fire under control time, we expected to have more responding personnel to deal with it. From the results, I find out that the result indeed captures the hypotheses. The results suggest a multiple linear regression relationship between the under control time, extent of fire, and number of responding personnel. Taken together, I used the practical rationale and p-value significance cut-off to find out the estimated multiple linear regression model for this report. That is $\hat{Y}_i = 19.81 + 0.19X_{i,1} + 8.51X_{i,A} + 0.15X_{i,1} \times X_{i,A}$, where $\hat{Y}_i$ is expected number of responding personnel, $X_{i,1}$ is the under control time and $X_{i,A}$ is the extent of fire. Practically, when we find out that the fire is not confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.3392475. When we find out that the fire is confined to the object of origin, then until the fire is controlled, every minute, the average number of responding personnel expect to increase by 0.1899735. Moreover, I have checked the assumptions of the linear model. The constant variance assumption may be violated in the model. Therefore, precision in our estimate would be less, and the hypothesis test for the coefficients in the linear regression model may be less reliable. In addition, From figure 4, we can see that we expect to have approximately 23 responding personnel present most of the time during a fire incident. Besides, we found that the number of responding personnel is right-skewed. However, we can apply CLT as the sample size is large enough. The distribution of the sample mean would converge to a Normal distribution.
Besides, I have a detailed data collection and cleaning process with variable descriptions in the data section.

Also, I want to emphasize several critical results in the data section. Table 2 found that the fire under control time has a relatively more considerable fluctuation and spread. According to table 3 and figure 1, we know that the fire that is not confined to the object of origin accounts for the most significant proportion(54.31%), which is around 1.2 times of the fire confined to the object of origin(45.69%). Hence, fire not confined to the object of origin occur more frequently than fire confined to the object of origin. Furthermore, from figure 2, we would expect that most of the time, regardless the fire is or is not confined to the object of origin, we would have nearly the same number of responding personnel. However, for the fire not confined to the object of origin, we would expect significant fluctuation for the number of responding personnel. Most likely, we would expect more responding personnel. From figure 3, we can see a different trend for the extent of fire. For the fire not confined to object of origin, the number of responding personnel is greater for the same under control time than confined to object of origin. It could form a strong positive linear relationship between fire under control time, extent of fire and the number of responding personnel.
Practically, we could apply these results to the fire scene. These results could help the commanders to make decisions on the plan depending on whether the fire is confined to the object of origin. If the fire is not confined to the object of origin, we would expect more responding personnel needed for every minute until the fire is controlled. Thus, I would suggest that once the commanders find out the fire is not confined to the origin. We should rapidly survey the scope of the fire, working out an effective plan to control the fire. For the witness or the people present at the fire scene, we should try our best to prevent contact between the object of origin and combustible material, trying to confine the fire at the origin rather than spread out elsewhere.
Overall, this report uses the multiple linear regression model to help the fire commander or the people at the fire scene to make decisions depending on the scientific model by quantifying the relationship, rather than just empirical reasoning.

Weaknesses

This study may only be generalizable to the fire incidents in Toronto since the population and data in this study are only the fire incidents in Toronto.
There may have been some other confounding variables in this report's multiple linear regression model, and this report's multiple linear regression model may not be the most significant.
Moreover, the constant variance assumption may be violated in the model. Therefore, precision in our estimate would be less, and the hypothesis test for the coefficients in the linear regression model may be less reliable.

Next Steps

It could be interesting to do some research regarding the confounding variables that are not listed in the original dataset. However, I cannot discuss it here since I do not have the data regarding the other variables related to a fire incident in this dataset.

Discussion

Throughout the report, my research question is to study what are the variables that could have a relationship with the number of responding personnel. Firstly, I investigated the numerical and graphical summaries for extent of fire, under control time and number of responding personnel in the data section. Then, I came up with the method I used to study the relationship, and I found that the extent of fire and fire under control time could have a linear relationship with number of responding personnel by using practical rationale and p-value significance cut-off. The methodologies I used were introduced in the method section and all corresponding results are in the result section.

Bibliography

All analysis for this report was programmed using `R version 4.0.3`.

1. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: October 25, 2021)
2. City of Toronto. (2020, December 16). *Fire services.* City of Toronto. Retrieved October 25, 2021, from https://www.toronto.ca/city-government/accountability-operations-customer-service/city-administration/staff-directory-divisions-and-customer-service/fire-services/.
3. David Robinson, Alex Hayes and Simon Couch (2021). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.9. https://CRAN.R-project.org/package=broom
4. Daignault, K. (2021, October). *Module 3 Asynchronous Material: Assumptions and Properties of Estimators.* Lecture. Retrieved Oct 25, 2021, from https://q.utoronto.ca/courses/236123/pages/module-3-asynchronous-material?module_item_id=3011149
5. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.
6. Friebe, M., &; Lasiuk, J. (2003). *A history of the Toronto fire services, 1874-2002.* Toronto Fire Services.
7. Ganti, A. (2021, September 21). *What is the Central Limit Theorem (CLT)?* Investopedia. Retrieved October 25, 2021, from https://www.investopedia.com/terms/c/central_limit_theorem.asp.
8. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: October 25, 2021)
9. Open data dataset. (2021). *City of Toronto Open Data Portal.* Retrieved October 23, 2021, from https://open.toronto.ca/dataset/fire-incidents/.
10. *Toronto fire services master fire plan.* (2019). Retrieved October 24, 2021, from https://www.toronto.ca/wp-content/uploads/2017/08/96f4-2015-2019_Master_Fire_Plan_WEB.pdf.