MATH 245

Prof. Laura Chihara

Ziyang Gao & Sarah Leong-Fern

Predicting Casualties of Terrorist Attacks

Introduction

With both casualties and economic loss caused by terrorism reaching an all-time high in the last year, the study of terrorism is more important than ever. In order to minimize casualties of terrorism, it is useful for policymakers and law enforcement to understand factors that play significant roles in the number of people killed or injured in an attack. So that whenever another terrorist attack happens, governments could react smoothly upon the situation and get prepared according to the estimated casualty derived by our model. On the other hand, governments or related organizations could look specifically into those factors that are significantly influencing number of casualty, such as entities and regions, in order to make effective arrangements of limited safety protection resources.

There have been multiple studies previously addressing this question. For example, the FPRI Center for the Study of Terrorism conducted a fact-based analysis of the effectiveness of actual and potential uses of terrorism as a tactic by adversaries of the United States and its allies; The *Incidents of Mass Casualty Terrorism* compiled by Wm. Robert Johnston whom did a careful and thorough work of concluding casualty for every single terrorist attack for the past century. However, there are not many sophisticated scholarly works dedicated to predicting the civilian casualty counts of terrorist attacks based on multiple related factors. So we believe that

our project is reasonably interesting and meaningful, even though we only have a few variables for our model.

To study this, we first define terrorism. Among the existing arguments in scholarly circles over definition of terrorism, there are a few things that are generally agreed upon:

1. It is a violent act.
2. It is committed with the purpose of terrorizing people. For example, if a terrorist assassinates a politician, the goal of the attack is cause fear among other politicians, not just the death of that specific person.
3. It is committed for political or ideological reasons. For example, if a mobster attacks a local business with the aim of causing fear, but their goal is only to extract money from the local community, it doesn't count as terrorism.
4. The perpetrators can't be working for a state. Not all scholars agree on this last point, but considering the dataset that we're using, this is an important part to include.

There have been studies on the subject of terrorist effectiveness before, but experts have not come to any general agreement. There is much evidence to suggest that the tactics and casualties of terrorist attacks are different depending on what region of the world the attack occurs in. For example, in 2014, the total number of terrorist assassinations committed in the Middle East was several times larger than the total number of terrorist assassinations elsewhere. In our study, we will try to determine what factors make a terrorist attack likely to have higher casualties.

Methods

For this study, we will be examining a subset of the Global Terrorism Dataset. We got the files for this dataset from a class, Methods of Political Research, which Sarah previously took, although she did not use this data for any research in that class. In this dataset, each observation is one terrorist attack. We have chosen to all the terrorist attacks from 1970 to 1997 which have a completely trackable data of all the variables we are interested in, which means we intentionally omitted the those attacks that are only partially recorded from our study. Here are some logistic behind our method of selecting data:

1. Terrorist attacks happen independently in each case with basically no correlations with each other, so the sequential data we use across years wouldn't cause a problem of independence assumption.

2. Making up the lost part of data create bias to our final model, and considering we have a relatively large data set (3000 observations), the omitted data wouldn't cause a problem of out degrees of freedom.

3. Since each terrorist attack is presumably independent of other terrorist attack and the losing data records happens completely at random, omitting a small portion of fragmentary data from a huge data pool wouldn't introduce much bias.

In this study, we will use regression to predict the response variable of how many people were killed in the terrorist attack. The explanatory variables are: region of the world, type of target, log of the number of terrorists in the group, type of victim (e.g. business), type of incident (e.g. assassination), and types of weapons used in the attack. The response variable is the number of people killed in the attack.

Results: Summary of Variables

1. Numerical variables: number of people killed (response variable) & number of terrorists involved in attack (explanatory variable)

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Stan Dev
# people killed	0	0	1	1.53	1	115	4.74
# terrorists	1	2	3	5.97	4	999	29.53

Table 1: Summary statistics for 3263 observations

From the table above we found that both of the variables abruptly increase a lot from 3rd quartile to maximum value. Plus, figures below (Fig.1, 2, 3) are our response variable (number of casualties) against number of explanatory variable (number of terrorists). As we taking out outliers, more outliers appear in our graph, so we decided to log transform the number of terrorist variable (as in Fig.4).

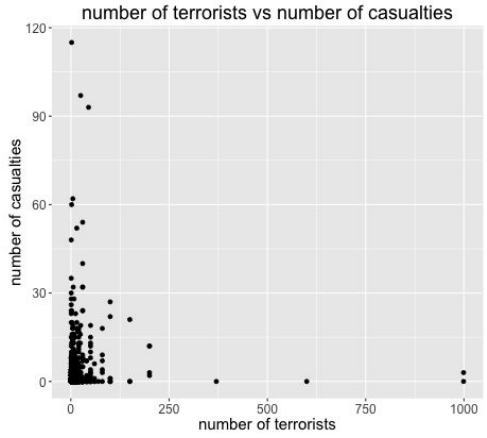

Fig.1

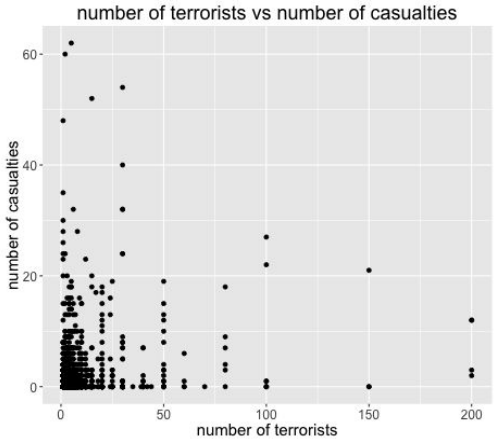

Fig.2

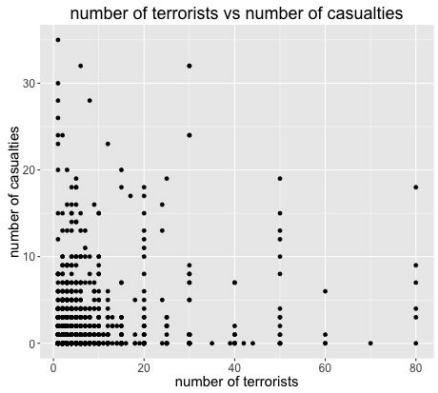

Fig.3

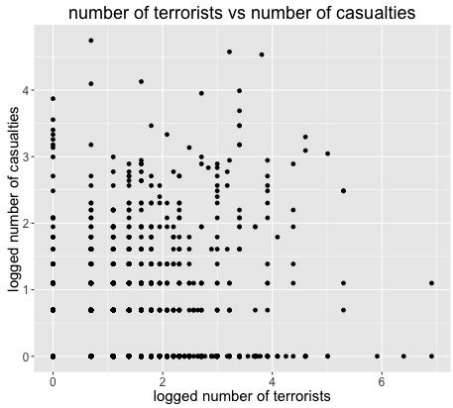

Fig.4

We decided to choose Poisson regression as our type of model rather than performing a regression of E[log(Y)], because our response variable is a count and doesn't have constant variance.

2. Categorical explanatory variable: region of the world attack took place in

Region	North America	Latin America	Europe	Middle East/North Africa	Sub-Saharan Africa	Asia
# of attacks	41	1072	690	611	171	755

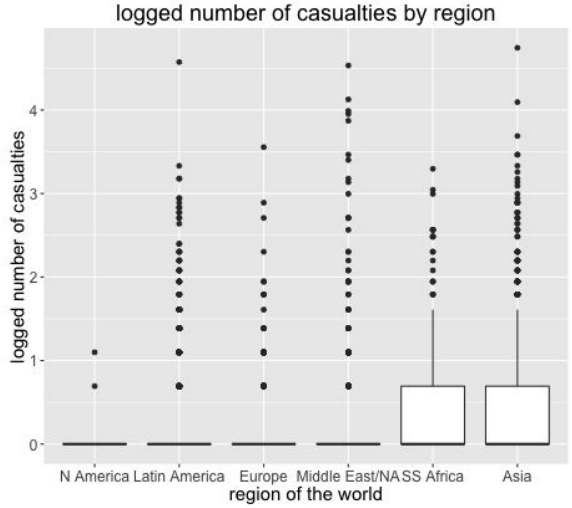

Fig.5 A boxplot for each region of the world showing logged number of casualties for that region

3. Categorical explanatory variable: type of attack

Type of attack	Assassination	Bombing	Facility Attack	Hijacking	Kidnapping
Number of attacks	1589	312	1076	93	192

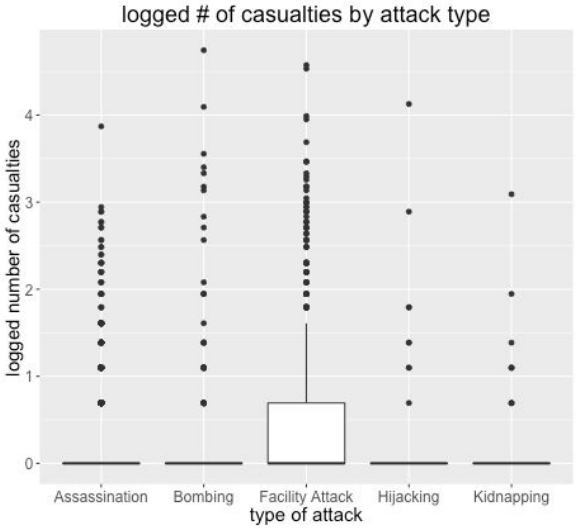

Fig. 6 A boxplot for each attack type showing logged number of casualties for that attack type

4. Categorical explanatory variable: weapon of attack

Type of weapon	Explosives	Firearms	Fire or firebomb	Knives & sharp objects	Chemical agent	Other
Number of attacks	421	2523	97	159	4	58

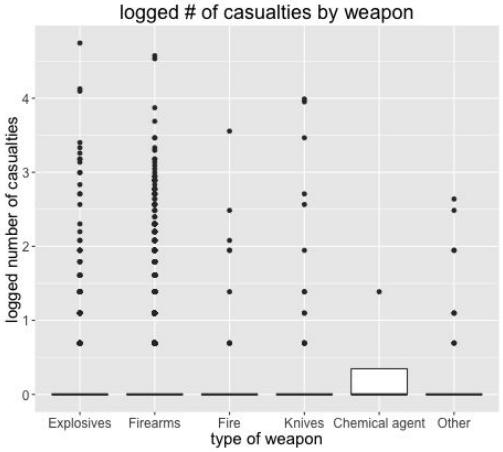

Fig. 7 A boxplot for each type of weapon showing logged number of casualties for that weapon

5. Categorical explanatory variable: type of target

Type of target	# of attacks
Diplomat	118
Police/Military	786
Other	316
Unknown	216
Government	458
Political Parties	358
Media	157
Business	619
Transportation	172
Utilities	15
International	39

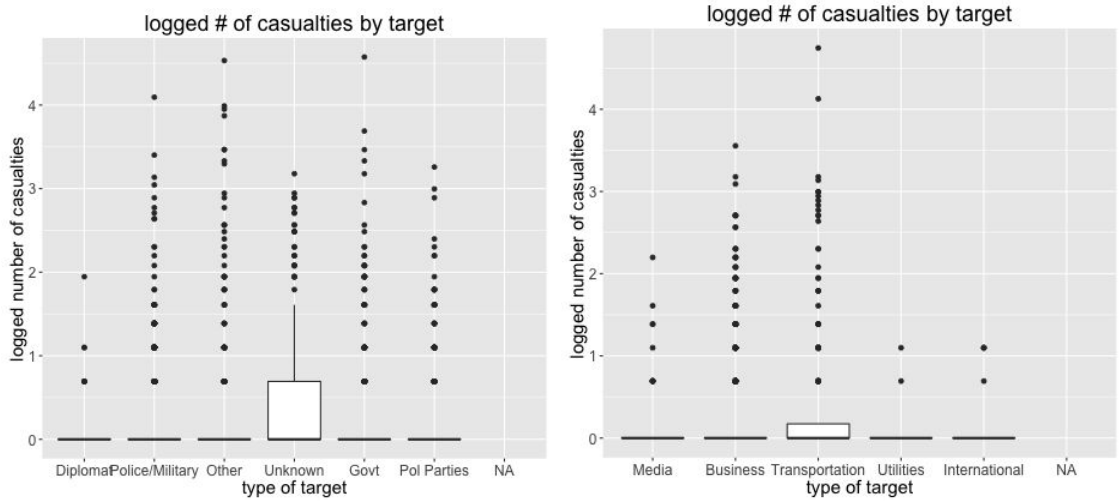

Fig. 8 and Fig 9. A boxplot for each type of target showing logged number of casualties for that target

Results: Model

We know from our initial exploration of the response variable that we want to use a Poisson regression. It fulfills all the assumptions: the response variable is made of counts, the observations are independent, and the variance of y is not the same for each x. We also checked diagnostic plots to make sure there were no major irreguarities with the model (see Appendix.)

The model is as follows:

E[Y] = μ = number of people killed in terrorist attack

Log(μ) = β0 + β1(log number of terrorists involved in the attack) + β2(Attack took place in sub-Saharan Africa) + β3(Attack took place in the Middle East) + β4(Attack took place in Asia) + β5(attack was on a facility or infrastructure) + β6(attack was a kidnapping) + β7(weapon of attack was firearms) + β8(weapon of attack was fire or a firebomb) + β9(weapon of attack was knives) + β10(weapon of attack was not a common weapon) + β11(interaction term for log number of terrorists and weapon of attack was fire) + β12(interaction term for log number of terrorists and weapon of attack was knives) + β13(victim of attack was police or military) + β14(victim of attack was not a common victim type) + β15(intended victim remains unknown) + β16(victim of attack was a government) + β17(victim of attack was a political party) + β18(victim of attack was a business) + β19(victim of attack was a type of transportation)

	name	estimate	Std error	p	95% CI for coefficient
β0	intercept	-1.05640	0.22807	3.77e-06	(-0.609, -1.503)
β1	logperps	0.39057	0.03669	< 2e-16	(0.317, 0.462)
β2	SSAfrica	0.40003	0.15631	0.01054	(0.706, 0.937)
β3	MiddleEast	0.64552	0.10610	1.31e-09	(0.438, 0.853)
β4	Asia	0.72563	0.09110	2.26e-15	(0.547, 0.904)
β5	FacilityAttack	0.21491	0.08649	0.01301	(0.045, 0.384)
β6	Kidnapping	-1.44474	0.32285	7.91e-06	(-2.078, -0.812)
β7	Firearms	-0.27269	0.10657	0.01055	(-0.482, -0.064)
β8	Fire	-0.24576	0.42343	0.56168	(-1.076, 0.584)

β9	Knives	-1.42632	0.34476	3.61e-05	(-2.102, -0.751)
β10	WeaponOther	-0.81984	0.35692	0.02168	(-0.120, -1.152)
β11	logperps*Fire	-0.63975	0.32332	0.04793	(-1.273, -0.006)
β12	logperps*Knives	0.44700	0.11438	9.50e-05	(0.223, 0.671)
β13	PoliceMilitary	0.71294	0.21745	0.00105	(0.287, 1.139)
β14	VictimOther	1.37811	0.22450	9.34e-10	(0.938, 1.818)
β15	Unknown	1.29936	0.23579	3.85e-08	(0.837, 1.761)
β16	Govt	0.89294	0.22484	7.30e-05	(0.452, 1.334)
β17	PolParties	0.88053	0.23724	0.00021	(0.416, 1.346)
β18	Business	0.64722	0.22454	0.00397	(0.207, 1.087)
β19	Transportation	1.46330	0.23516	5.52e-10	(1.002, 1.924)

Because we have so many explanatory variables, we will take only the numerical explanatory variable and one of the categorical explanatory variables to show the relationships between the 95% confidence intervals for the slopes and their interpretations.

We are 95% sure that for every time the number of terrorists involved in an attack is doubled, the median number of resulting casualties will be between 90% and 78.7% smaller.

We are 95% sure that every attack that occurs in sub-Saharan Africa will have between 103% and 155% more casualties than if it had not occurred in sub-Saharan Africa.

Discussion

Our main findings are:

1. The region that the attack took place in, the type of victim the attack was targeted at, and the type of weapon used by the attacker are variables that influence the number of casualties significantly;
2. There is a significant interaction between the number of attackers and the type of attack used;
3. The more attackers there are, the less casualties there are.

These all are as we predicted, except for the influence of the number of attackers. We thought that more attackers could inflict more damage. One way to explain this relationship is by looking at the interaction term between number of terrorists and weapon type. The weapon type knives has a positive interaction with number of terrorists, meaning that knife attacks are far more effective with many terrorists. Perhaps the most effective attacks, however, rely on only a few terrorists carrying out a bombing or chemical attack. From that, we can conclude that law enforcement should put as much effort into lone wolves as they do into complex terrorist organizations, as only one, two, or a few terrorists can have just as much of an impact in one attack as several.

Of course, this is only for the years 1970-1997. Since then, there have been several developments in terrorism, including a move away from kidnapping and hijacking and towards bombing as a choice of tactics. Additionally, there is more suicide terrorism than before. An interesting follow-up study would use the GTD data for a more recent year and compare the findings from the two studies.

Appendix

1. Decision about keeping in the insignificant term logperps and its interaction term: in the summary output the p-values for both the variable logperps and the Fire were not significant; however, in the drop in the deviance test, we got a very small p value. So we decided to keep those terms in. Below is the test process for logperps, same as for Fire.

```
> anova(terror.glm6, terror.glm5)
Analysis of Deviance Table
Model 1: nkill ~ region + FacilityAttack + Hijacking + Kidnapping + Firearms +
    Fire + Knives + WeaponOther + PoliceMilitary + VictimOther +
    Unknown + Govt + PolParties + Business + Transportation
Model 2: nkill ~ logperps + region + FacilityAttack + Hijacking + Kidnapping +
    Firearms + Fire + Knives + WeaponOther + logperps * Firearms +
    logperps * Fire + logperps * Knives + logperps * WeaponOther +
    PoliceMilitary + VictimOther + Unknown + Govt + PolParties +
    Business + Transportation
  Resid. Df Resid. Dev Df Deviance
      1      3235      11309
      2      3230      10062  5   1247.8
> 1-pchisq(1247.8, 5)
[1] 0
```

2. The VIF table below shows multicollinearity of each of the variables, including some of the interaction terms that we think are important. It turns out that none of the VIF value is larger then 10, which means that none of the variables is highly correlated with others.

> vif(terror.glm12)

logperps	SSAfrica	MiddleEast	Asia	FacilityAttack	Kidnapping
1.502378	1.176899	1.326237	1.294689	1.292839	1.036171
Firearms	Fire	Knives	WeaponOther	PoliceMilitary	VictimOther
1.436856	1.949064	3.778931	1.080513	5.512300	4.629090
Unknown	Govt	PolParties	Business	Transportation	logperps:Fire
3.508624	4.412932	3.404371	4.434795	3.478006	1.863282
logperps:Knives					
3.840929					

3. Cook's Distance: none of the Cook's distance value is larger than 1, so none of the observations have a dramatic effect to the model when deleted from the data pool.

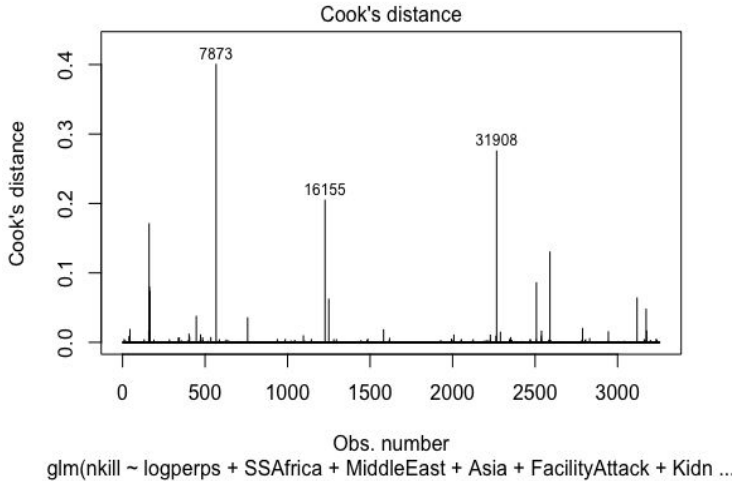

Fig.1 Cook's distance

4. Residuals Plot: Since a large amount of our predicted means are less than 5, our distributions of residuals is not approximately normal (which makes sense). So in this case, we'll just use the residual plot to spot outliers. From the residual plot, point 16948, 31908 and 39392 are the most "outlied", however, taking them out resulting in more outliers appear so we'll just keep them in the model since we have 3262 data points in total, three of them won't make a hugh difference in the the overall interpretation of our final model.

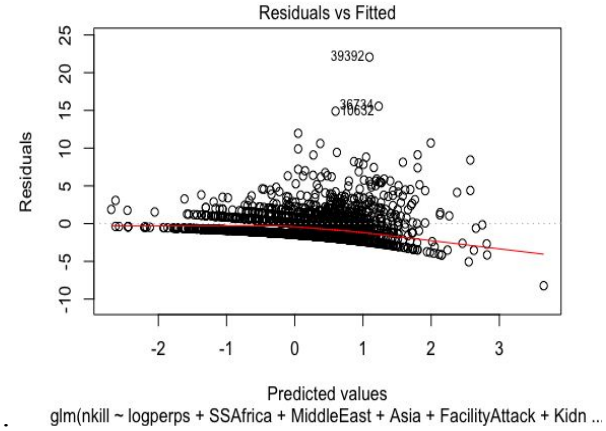

Fig.2 First time taking out outliers

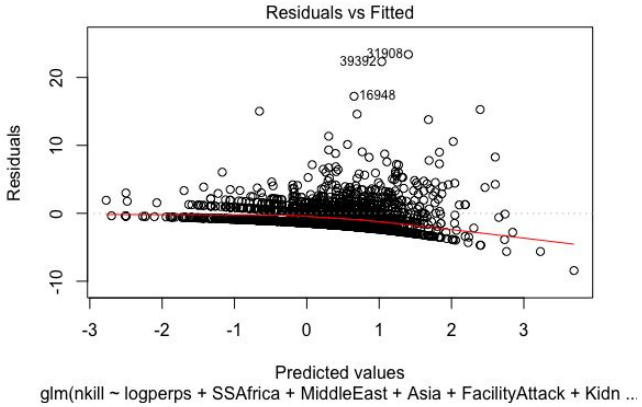

Fig.3 Second time taking out outliers

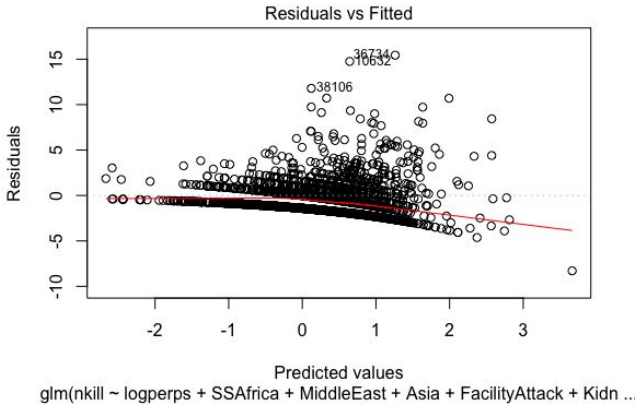

Fig.4 Third time taking out outliers