Global Terrorism

OMIS 645 – Applied Business Analytics in SAS

Dr. Kishen Iyengar

December, 5th 2018

Team Members

Anitha Nagaraju, Z1829732

Petia Stoyanova-Pouhaleva, Z1834315

Rojina Thapa, Z1834019

Sajeed Shaikh, Z1720772

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Countries affected by global terrorism tend to suffer and these types of incidents tend to have high impact on their economy. Since it has a global affect, the main purpose of this report is to measure and understand various attributes that are impacting global terrorism. The dataset used in this report consists of 26 variables and 5,962 observations.

Our team started this project asking interesting questions about our dataset. To answer them, we ran multiple tests such as linear/logistic regression, moderation analysis, and test of independence. In addition, with all our models being significant, we came up with an abundance of interesting findings about the global terrorism dataset. In our project, we used SAS University Edition to extract and interpret data by applying various statistical methods.

DATA SOURCE

We obtained our Global Terrorism dataset from Kaggle, an online community owned by Google, where various data scientists and machine learners post datasets. Before considering our dataset, we analyzed a few options after which we decided to select one that was a current global issue. Our Global Terrorism dataset is rich in the number of variables and observations that it contains. We tried looking for other sources of data to add onto out current dataset, but many of the variables that we found in the new datasets were very similar to our existing ones. Hence, we decided not to integrate them with our current data to avoid multicollinearity among similar variables.

DATA CLEANING

Cleaning columns

Our dataset initially consisted of 135 variables, some of the columns contained unique variables and some even contained sentences. In order to retain only the meaning full values, we removed the variables we felt were not useful for further analysis. After removing them, we were left with 26 variables.

Cleaning rows

Since the file size was too large to run in SAS, the number of rows was reduced from 182,000 to rows from the last 10 years (Jan. 2008 – Dec. 2017), to make it more relevant. Before our initial analysis, it was observed that there were a lot of missing values in our data-set in the variables that were to be used for statistical analysis. In order to handle these missing values, we deleted the rows that contained these missing values and our data set now resulted in 5962 observations.

VARIABLE DESCRIPTIONS

Variable Name	Values	Description	Type
Year	2007-2017	Year the attack took place	Numerical
Month	1-12	Month the attack took place	Numerical
Day	1-31	Day the attack took place	Numerical
Extended 24hrs	Yes/No (1/0)	If the attack is extended more than 24 hours	Binary Categorical
Country	95 Countries	Country the attack took place	Categorical
Region	Africa, Asia, Australia, Europe, North America, South America	Region in which the terrorist attack took place	Categorical
Multiple Attacks	Yes/No (1/0)	If there are multiple attacks	Binary Categorical

Success or Not	Yes/No (1/0)	If the attack is successful	Binary Categorical
Suicide Attack	Yes/No (1/0)	If the attack is suicidal	Binary Categorical
Attack Type	Armed Assault, Assassination, Bombing, Facility Attack, Hijacking, Hostage Taking, Unarmed Assault	Type of attack that took place	Categorical
Target Type	Abortion, Airports Business, Educational, Food/Water, Government, Media, Maritime, Military, NGO, Other, Police, Private Citizens, Religious, Telecommunication, Terrorists, Transportation, Utilities, Violent Political Party	The targeted cause during the attack	Categorical
Targeted Nationality	United States, Uzbekistan, Venezuela, Vietnam, Virgin Islands (US), Yemen, Zimbabwe, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Turkey, Uganda, Ukraine, Russia, Rwanda, Saudi Arabia, Senegal, Serbia-Montenegro, Slovak Republic, Somalia, South Africa, South Korea, South Sudan, Spain, Sri Lanka, St. Lucia, Sudan, Sweden, Pakistan, Philippines, Poland, Qatar, Nigeria, Niger, Northern Ireland, Nepal, Nederland, New Zealand, Libya, Madagascar, Malaysia, Maldives, Mali, Mexico, Morocco etc.	The nationality targeted in the attack	Categorical
Gun Used	Yes/No (1/0)	If there was a usage of a gun or not	Categorical
Number of Terrorists	0-3000	Number of terrorists involved in the attack	Numerical

Number of Terrorists Captured	0-63	Number of terrorists captured in the attack	Numerical
Weapon Type1	Chemical, Explosives, Firearms, Incendiary, Melee	Type of weapon used during the attack	Categorical
Weapon Type2	Arson/Fire, Semi/Automatic Rifle, Blunt Object, Dynamite, Explosive, Gasoline, Grenade, Handgun, Hands, Knife, Landmine, Letter Bomb, Molotov, Other, Pipe Bomb, Poisoning, Pressure, Projectile, Remote, Rifle, Rope, Sticky Bomb, Suicide, Time Fuse, Vehicle	Type of weapon used during the attack	Categorical
People Killed	1-588	Number of people killed in the attack	Numerical
US People Killed	1-44	Number of US people killed in the attack	Numerical
Number of Terrorists Killed	1-136	Number of terrorists killed in the attack	Numerical
Number of people wounded	1-491	Number of people wounded in the attack	Numerical
Number of people wounded US	1-57	Number of people wounded US in the attack	Numerical
Number of Terrorists Wounded	1-50	Number of Terrorists Wounded in the attack	Numerical
Property	0/1	If property was damaged or not	Binary Categorical
Property Value	1-350000000	Value of Property that was damaged	Numerical
Kid Held Hostage	Yes/No (1/0)	If kids are held hostage in the attack	Binary Categorical

STATISTICAL ANALYSIS OF SELECTED DATASET

QUESTION 1

What are we trying to answer?

We are trying to figure out the relationship between the number of people being killed during a terrorist attack and the various attribute of a terrorist attack such as the weapon involved, the type of attack, the region, etc.

Dependent Variables

We can clearly see the dependent variable here is the number of people killed during a terrorist attack, we picked this as our dependent variable because we thought it would be interesting to know what factors of a terrorist attack contribute to killing more people when compared to others.

Independent Variables

To pick the independent variables for this analysis, we examined various variables in the Global Terrorism data to find those that we thought would affect the number of people killed in a terrorist attack. In doing so, we identified the below variables and ran an analysis to find that one of the variables – Extended 24 hrs., which tells us if the attack lasted more than 24 hours, was not a strong predictor of the number of people killed.

Hence, to make the analysis for efficient we removed the variable from our model and ran the analysis again with the remaining variables. From the output obtained, we could clearly see that there were 5 factors that affected the number of people killed

Note: The dataset initially had 12 different regions which we classified into the below 6 continents for this specific analysis:

South America – South America

Africa – Middle east & North Africa, Sub-Sharan Africa

Asia – Central Asia, East Asia, South Asia, Southeast Asia

Australia – Australasia & Oceania

Europe – Western Europe, Eastern Europe

North America – Central America & Caribbean, North America

South America – South America

Analysis Model Used - Zero Inflated Negative Binomial

Realizing that our independent variable is a count variable and the data set is inflated with many zeros we analyze this question using zero-inflated negative binomial distribution. This distribution is generally used when there are a large number of zeros in the data taken into consideration. As an alternative, we tried performing log transformations to remove the zeroes in the data, resulting in undefined values for log (0). Another possible solution was to remove all the zeros, but doing do would affect the veracity of the data taken for consideration.

We also noticed that the variance of the data is larger than the mean even when all the 0's were removed from the dataset. Hence, due to the large number of 0's in the variables considered, we used the Zero-inflated Negative Binomial Distribution to analyze the data for this question.

Interpretations

With the given output in **Figure 1.3** we are trying to understand which among the independent variables are related to the dependent variable, and we further explore the forms of these relationships by looking at the parameter estimates. From the low p-value seen in **Figure 1.3** we can say that this model is significant.

From the Analysis of Maximum Likelihood Parameter Estimates Table, we come up with the Regression line shown below.

Number of people killed = -2.3786 +0.0079***NumberofTerrorists -** 0.374***Chemicals** + 1.9406***Explosives** +1.0290***Firearms** - 0.4263***Incendiary** + 0.9888***Africa** + 0.7378***Asia** - 15.7804***Australia** + 0.6691***Europe** - 0.1902***North America** + 1.13148***ArmedAssault** +0.8563***Assassination** + 1.4391***Bombing** - 0.661***InfrastAttack** +1.2790***Hijacking** + 1.95***HostageTaking -** 0.1167***Gun_Used**

Interesting Findings (Figure 1.3)

- ✓ We would normally assume that people killed in a terrorist attack due to incendiaries would be a lot more than deaths occurring due to Melee weapons. Our findings contradict this assumption and proves the opposite. When compared to Melee, the number of people killed by incendiaries is 0.4263 times less.
- ✓ Compared to Melee, the number of people killed when using explosives is almost twice as much.
- ✓ Compared to Melee, the number of people killed when using explosives is 1.9406 times as much.
- ✓ With a unit increase in the number of terrorists involved in a terrorist attack the number of people killed increases by 0.0079.
- ✓ As expected, when compared to no guns being used during a terrorist attack, there were more people killed when guns were used.
- ✓ When compared to an unarmed attack there were twice the amount of people killed during a hostage taking.
- ✓ Compared to an unarmed assault, attacks involving armed assault and attacks involving hijacking killed 1.3 times as many people. This was similar to a bombing or explosion attack, which killed 1.4 times as many people.

QUESTION 2

What are we trying to answer?

This question focuses on regions of the world having a significant impact on whether a terrorist attack is a suicidal attack or not.

Independent Variables

The independent categorical variable used in this analysis is region. There are 12 regions that are observed: Australasia & Oceania, Central America & Caribbean, Central Asia, East Asia, Eastern Europe, Middle East & North Africa, North America, South America, South Asia, Southeast Asia, Sub-Saharan Africa, and Western Europe. Because region (12) far less categories than countries (95), it would be favorable to perform a test that would yield significant results.

Dependent Variables

The dependent categorical variable is whether a terrorist attack was suicidal or not (0 – No, 1 - Yes). The level of interest of this categorical variable is 1 (explained in interesting findings). This variable was chosen because according to The Washington Post, “The suicide attacks meet our [readers] definition of terrorism”

Analysis Model Used - Test of Independence

A Test of Independence was used in this question because we are dealing with two categorical variables. We want to see if they impact one or the other. This tests whether the variables are dependent or not. The hypotheses that are testing with degrees of freedom are:

H_0: Region and Suicidal Attack are independent

H_1: Region and Suicidal Attack are dependent

Interpretations

In **Figure 2.3**, the chi-square p-value is less than 0.05, meaning that it is a significant model

In **Figure 2.3**, the degrees of freedom for this test are 11. df = (12-1)*(2-1) = 11

The model is significant, NULL hypothesis is rejected. Thus, there's evidence the variables are dependent.

Interesting Findings

- ✓ Based on the distribution of region by suicide attack plot (**Figure 2.2**), the regions with the top three largest distributions of suicidal attacks are Sub-Saharan Africa, South Asia, and Middle East & North Africa.
- ✓ Based on the distribution of region by suicide attack plot (**Figure 2.2**), the region with the largest distribution of non-suicidal attacks is Southeast Asia.
- ✓ Readers would normally assume that the Middle-East & North African region would yield the largest distribution of suicidal attacks because they believe it to be an analytical strategy that is widely practiced in that region. This analysis proves to be misleading with South Asia and Sub-Saharan Africa both having almost 2/3^{rds} the amount of suicidal attacks as the Middle East & North African region.
- ✓ The Middle East & North Africa has more suicidal attacks than non-suicidal attacks (**Figure 2.1**).

QUESTION 3

What are we trying to answer?

We are trying to answer if Number of terrorists, Weapon Type 1 and Extended 24 hours affect number of terrorists killed in the attack.

Dependent variable

The dependent variable here is the number of terrorists killed, which is a numerical variable. We chose number of terrorists killed because we wanted to see what attributes of attack affect the number of terrorists killed. We have many variables in our dataset which we tend to think affects the number of terrorists directly and in-directly, so we wanted to examine it against those variables.

Independent Variables

We have number of terrorists, weapon type 1 and attack extended 24 hours as independent variables here. Number of terrorists is the numerical variable. Weapon Type 1 is the categorical variable with values chemical, explosives, firearms, incendiary and melee. Attack extended more than 24v hours is categorical variable with values 0 and 1, 0 being no and 1 being yes. We had many variables which impact the number of terrorists killed. Out of those variables we chose these. We chose Weapon Type 1 and not Weapon Type 2 because Weapon Type 2 has many values and Binary Logistic Regression does not handle categorical value with many values well.

Analysis Model Used – Multiple Linear Regression

H_0: The model is not significant

H_1: The model is significant

Since the dependent variable is numerical, we use Linear Regression

Interpretation

In **Figure 3.1**, we can see that the p value is 0.0001 which is less than 0.05. So, the model is fit.

In **Figure 3.1**, Based on parameter estimates table, we came up with following regression line:

Number of Terrorists killed = 0.050455+0.003283***Number of Terrorists -** 0.128642***Chemical**

+0.625381***Explosives**-0.140641***Firearms** - 0.186205***Incendiary**+0.294197***Extended 24 Hrs**

Interesting Findings (Figure 3.1)

- ✓ Since the p-value of Number of Terrorists is less than 0.0001, we conclude that relationship between Number of terrorists and number of people killed significantly.
- ✓ Increase in Number of terrorists by 1 will increases the Number of people killed by 0.81906.
- ✓ Explosives increases the number of terrorists killed by 0.625381 more than Melee.
- ✓ Chemical, Firearms and Incendiary are not as significant as Melee.

- ✓ Extended 24 hours' p-value is 0.3558 so we conclude that it does not affect Number of Terrorists killed.

QUESTION 4

What are we trying to answer?

We want to know if suicide attacks, number of terrorists, multiple attack and gun used are predictor of attack being successful or not.

Dependent variable

The dependent variable here is Success, which is a binary categorical variable. Success has two values, 0 and 1, with 1 being successful and 0 being not successful. We wanted to see what attributes in our dataset contribute to the attack being successful. We chose success as our event of interest and measured it against many other independent variables.

Independent variables

We have Suicide Attacks, No. Of Terrorists, Multiple Attack (Categorical) and Gun Used as independent variables. Suicide Attack is the categorical variable here with values 0 and 1, 0 being no Suicide Attack and 1 being suicide attack. No. of terrorists is the numerical variable here. Multiple Attack is the categorical variable here with values 0 and 1, 0 being no multiple attack and 1 being multiple attack. Gun used is the categorical variable here with values 0 and 1, 0 being no gun used and 1 being gun used. We ran the test with several variables presented in the dataset But, those variables were not significant. Therefore, we chose the ones which were statistically significant.

Analysis Model Used - Binary Logistic Regression

The dependent variable used here is a binary categorical value, hence we perform binary logistic regression.

Interpretation

In ***Figure 4.1*** Since p-value is less than 0.0001, we conclude that the model is a good fit.

Interesting Findings (Figure 4.1)

- ✓ Multiple Attack is a significant predictor of success of attack.
- ✓ When there is no multiple attack, the log odds of success come down by 0.3295 less than when there is multiple attack.
- ✓ Suicide attack is a significant predictor of success of attack.
- ✓ When there is no suicide attack, the logs odds of success come down by 0.2121 less than when there is suicide attack.
- ✓ Gun used, and number of terrorists are a significant predictor of success of attack.
- ✓ When there is no gun used, the log odds of success come down by 0.5610 less than when there is gun used.

QUESTION 5

What are we trying to answer?

We are trying to figure out the relationship between the number of terrorists captured with the number of terrorists involved and attack type?

Dependent Variable

We chose the number of terrorists captured as dependent variable because we are interested in figuring out what attributes contribute the most to the number of terrorists captured.

Independent Variables

In order to pick the independent variables for this analysis, we examined various variables in the Global Terrorism data to find those that we thought would affect the number of terrorists captured during a terrorist attack.

Attack Type (Categorical) and Multiple Attack (Categorical)

We chose Attack type and Multiple Attacks as independent variable because we wanted to see if they have any impact on number of terrorists captured.

Analysis Model used - Multiple Linear Regression

H_0: The model is not significant

H_1: The model is significant

The dependent variable is numerical; therefore, we conduct a multiple linear regression analysis.

Interpretations (Figure 5.1)

Since p-value is less than 0.0001 in Analysis of Variance table, we conclude that the model is a good fit.

In Fig.5.1 we observed the parameter estimates and came up with the following regression line:

Number of Terrorists captured = 1.5267 - 0.2384***Multiple Attacks** – 0.9169***Armed Assault**

- 0.8731***Assassination** – 0.9977***Bombing** – 0.151***Facility** - 0.6105***Hijacking** – 0.521***Hostage Taking**

The interpretation of the intercept is that we expect Number of Terrorists captured to be equal to 1.5267 when attack types and multiple attacks are zeros. The expected change on average of the dependent variable Number of Terrorists captured for one unit of increase of the independent variable Multiple Attacks is -0.2384. In addition, the expected change on average of the dependent variable Number of Terrorists captured for one unit of increase of the independent variable Assassination is -0.8731. The relationship with all three statistically significant independent variables from attack type to the Number of Terrorists captured is negative.

Interesting Findings

- ✓ Four statistically significant independent variables.
- ✓ Negative relationship between the terrorists captured and the independent variables.

- ✓ When there are no multiple attacks, the number of terrorists captured decreases by 0.2384 less than when there are multiple attacks.
- ✓ Armed assault, assassination, bombings, hostage taking decreases the number of terrorists captured when compared to unarmed assaults and individual attacks.
- ✓ Compared to unarmed assault attack, an armed assault attack decreases the number of terrorists captured by 0.9169.
- ✓ Compared to Unarmed Assault, Assassination decreases the number of terrorists captured by 0.8731.
- ✓ Compared to Unarmed Assault, Bombing/Explosion decreases the number of terrorists captured by 0.9977.
- ✓ Compared to Unarmed Assault, Hostage Taking decreases the number of terrorists captured by 0.521.

QUESTION 6

What are we trying to answer?

We are interested to find out if the attack type and target type affect the multiple attacks in a terrorist attack.

Dependent variable

The dependent variable here is multiple attacks which is a categorical variable having the values (0 and 1). The former meaning that there are multiple attack and the latter meaning that there are no multiple attacks. The level of interest of the dependent variable is 1. The output is in reference to whether there are multiple terrorist attacks.

Independent variables

We examined various variables in the Global Terrorism data to find those that we thought would affect the multiple attacks. In doing so, we identified Attack Type: Categorical (1-9) and Target Type: Categorical to be effective.

Analysis Model Used - Binary Logistic Regression

H_0: Dependent variable is independent of the independent variables

H_1: Dependent variable is dependent of the independent variables

Since p-value is less than 0.0001 in Analysis of Variance table, we conclude that the model is a good fit.

Based on analysis of maximum likelihood estimates (**Figure 6.2**), we came up with following regression line:

Ln(p(MultipleAttacks=1)/(1-P(MultipleAttacks=1)) = -2.7682 + 0.3629***Armed Assault** - 0.9038 ***Assassination** – 1.4508***Bombing** + 1.2741**Facility** +1.3008***Hijacking** + 0.4643***Hostage Taking** - 0.0008***AbortionRelated -** 1.0316***Airports** – 0.0092***Business –** 0.0446***Education Institution** - 11.6392***Food/Water Supply** – 0.4936***Government (Diplomatic)** -0.3719***Government (General)** - 0.2964***Journalists** – 0.9841***Maritime** -0.2548***Military** – 0.3737***NGO** + 1.9497***Other** – 0.3763***Police** +0.1936***PrivateCitizens** + +0.1425***Religious Institution** + 1.546***Telecommunication** +0.1908***Terrorists**

We want to predict the probability of multiple attacks as a categorical dependent binary variable. In order to do so, we choose logistic regression analysis.

Interpretations

Model fit statistics:

The binary logit model is significant because the Likelihood ratio, Score, and Wald test have p-values less than .0001.

By using the logistic regression analysis, we want to predict the probability of multiple attacks as a categorical dependent binary variable. The categorical dependent variable can take only two possible types: 0 or 1. Furthermore, to evaluate the model fit, we use the -2log likelihood (-2 Log L) value. The fit between the model and the data is better with smaller values of -2 Log L.

From the Analysis of Maximum Likelihood Estimates, the statistically significant predictors for multiple attacks are:

- ✓ Telecommunication (target type) with p-value 0.0075
- ✓ Bombing/Explosion (attack type) with p-value 0.0156
- ✓ Facility/Infrastructure attack (attack type) with p-value 0.0365

Interesting Findings

- ✓ We would normally assume that from the seventeenth target types and seven attack types more would affect the multiple attacks. Our findings contradict this assumption and proves that only three of total twenty-four variables are statistically significant.
- ✓ The model predicted that the most statistically significant variable from target type is telecommunication.
- ✓ The most statistically significant variable from attack type are bombing/explosion and facility/infrastructure.

CONCLUSION

Based on the set of research questions that were answered via analysis of the modified dataset about global terrorism, we came up with interesting results that changed our preconceptions on how we view global terrorism. What we perceive as normal is based on what is viewed on television, blogs, social media, and news outlets. We see the beginnings, the events that occurred during the attack, the aftermath, and the impact it had on the people and surroundings. The main purpose of this report is to measure and have a better understanding of the relationship between attributes of global terrorism and its affect it has had on the world for the past 10 years.

For instance, in the last 10 years, the cause of deaths from a melee attack occurred more than incendiary attacks. The regions with the top three largest distributions of suicidal attacks are Sub-Saharan Africa, South Asia, and Middle East & North Africa. Number of terrorists wounded increases the number of people killed by 0.82. Multiple attacks, gun attacks, suicide attacks, and the number of terrorists involved are significant predictors of an attack being successful. Armed assault, assassination, bombings, hostage taking decreases the number of terrorists captured when compared to unarmed assaults and individual attacks. Bombing and infrastructure attack types as well as telecommunication target types significantly affect the chances of there being multiple attacks. Considering the limitations of this report, these findings can offer recommendations for future researches.

APPENDIX

Model Information		
Data Set	APLIB.GTREGION	
Distribution	Zero Inflated Negative Binomial	
Link Function	Log	
Dependent Variable	People_Killed	People Killed

Number of Observations Read	5962
Number of Observations Used	5962

Class Level Information		
Class	Levels	Values
Weapon_Type1	5	Chemical Explosives Firearms Incendiary Melee
Attack_Type	7	Armed Assault Assassination Bombing/Explosion Facility/Infrastructure Attack Hijacking Hostage Taking Unarmed Assault
Gun_Used	2	0 1
Region	6	Africa Asia Australia Europe North America South America

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		24460.4317	
Scaled Deviance		24460.4317	
Pearson Chi-Square	5943	15894.4242	2.6745
Scaled Pearson X2	5943	15894.4242	2.6745
Log Likelihood		-12230.2158	
Full Log Likelihood		-12230.2158	
AIC (smaller is better)		24500.4317	
AICC (smaller is better)		24500.5731	
BIC (smaller is better)		24634.2949	

Figure 1.1

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.7920	0.7099	-7.1835	-4.4005	66.56	<.0001

Figure 1.2

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.3786	0.5986	-3.5518	-1.2053	15.79	<.0001
Number_of_Terrorists		1	0.0079	0.0010	0.0060	0.0098	64.42	<.0001
Region	Africa	1	0.9888	0.4552	0.0965	1.8810	4.72	0.0299
Region	Asia	1	0.7378	0.4550	-0.1539	1.6295	2.63	<.0001
Region	Australia	1	-15.7804	9937.798	-19493.5	19461.95	0.00	0.9987
Region	Europe	1	0.6691	0.4612	-0.2349	1.5732	2.10	<.0001
Region	North America	1	-0.1902	0.4658	-1.1031	0.7227	0.17	<.0001
Region	South America	0	0.0000	0.0000	0.0000	0.0000	.	.
Weapon_Type1	Chemical	1	-0.1374	0.7716	-1.6496	1.3749	0.03	0.8587
Weapon_Type1	Explosives	1	1.9406	0.1301	1.6856	2.1956	222.52	<.0001
Weapon_Type1	Firearms	1	1.0290	0.1024	0.8283	1.2297	100.97	<.0001
Weapon_Type1	Incendiary	1	-0.4263	0.1922	-0.4614	0.2920	0.19	<.0001
Weapon_Type1	Melee	0	0.0000	0.0000	0.0000	0.0000	.	.
Attack_Type	Armed Assault	1	1.3148	0.3941	0.5423	2.0873	11.13	0.0009
Attack_Type	Assassination	1	0.8563	0.4030	0.0664	1.6462	4.51	0.0336
Attack_Type	Bombing/Explosion	1	1.4391	0.4062	0.6429	2.2353	12.55	0.0004
Attack_Type	Facility/Infrastructure Attack	1	-0.6610	0.4382	-1.5198	0.1978	2.28	<.0001
Attack_Type	Hijacking	1	1.2790	0.4826	0.3330	2.2249	7.02	<.0001
Attack_Type	Hostage Taking	1	1.9500	0.4112	1.1441	2.7559	22.49	<.0001
Attack_Type	Unarmed Assault	0	0.0000	0.0000	0.0000	0.0000	.	.
Gun_Used	0	1	-0.1167	0.0681	-0.2501	0.0167	2.94	<.0001
Gun_Used	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	2.1097	0.0505	2.0130	2.2111		

Figure 1.3

Frequency Expected

Table of Region by Suicide_Attack			
	Suicide_Attack(Suicide Attack)		
Region(Region)	0	1	Total
Australasia & Oceania	1 0.7035	0 0.2965	1
Central America & Caribbean	2 1.4069	0 0.5931	2
Central Asia	22 21.807	9 9.1929	31
East Asia	30 21.104	0 8.8963	30
Eastern Europe	215 196.97	65 83.033	280
Middle East & North Africa	766 1081.9	772 456.09	1538
North America	497 354.54	7 149.46	504
South America	21 14.773	0 6.2274	21
South Asia	663 804.05	480 338.95	1143
Southeast Asia	1339 947.55	8 399.45	1347
Sub-Saharan Africa	473 629.59	422 265.41	895
Western Europe	165 119.59	5 50.413	170
Total	4194	1768	5962

Figure 2.1

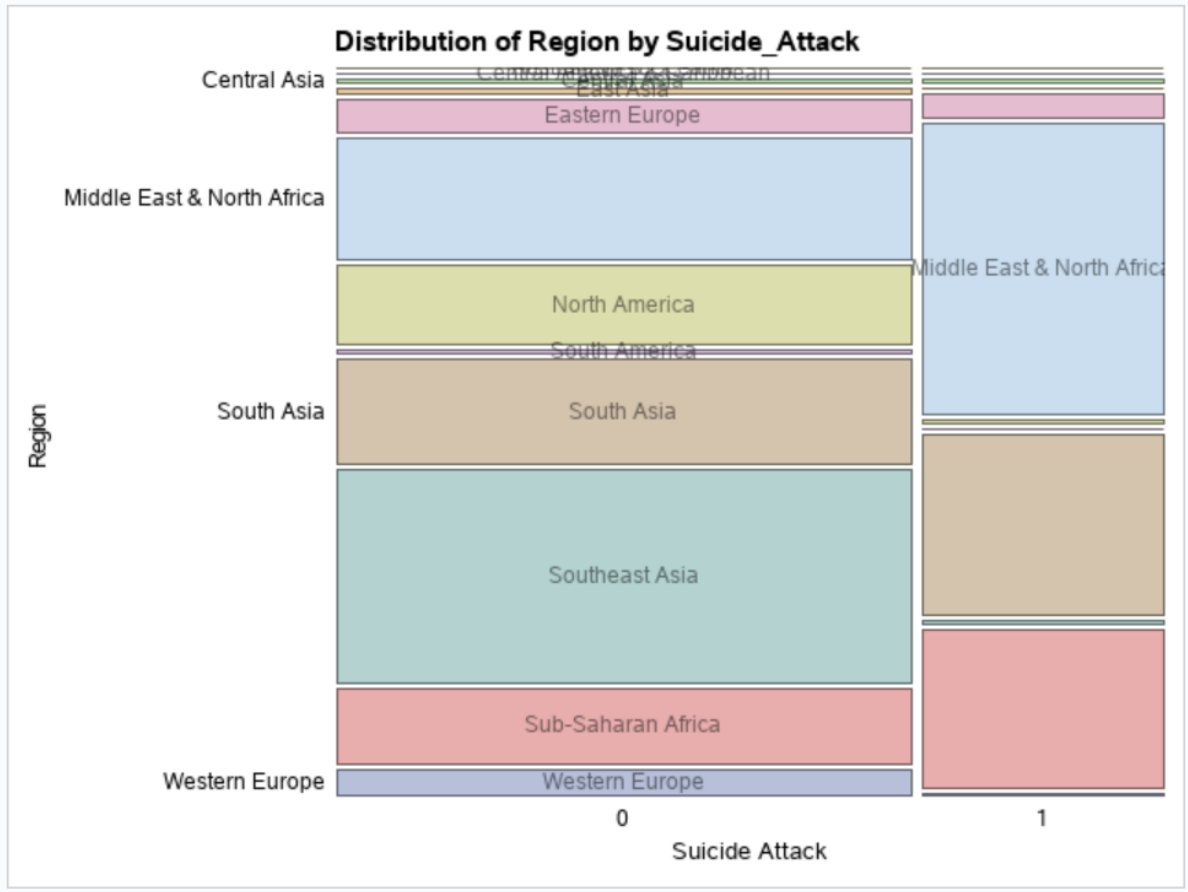
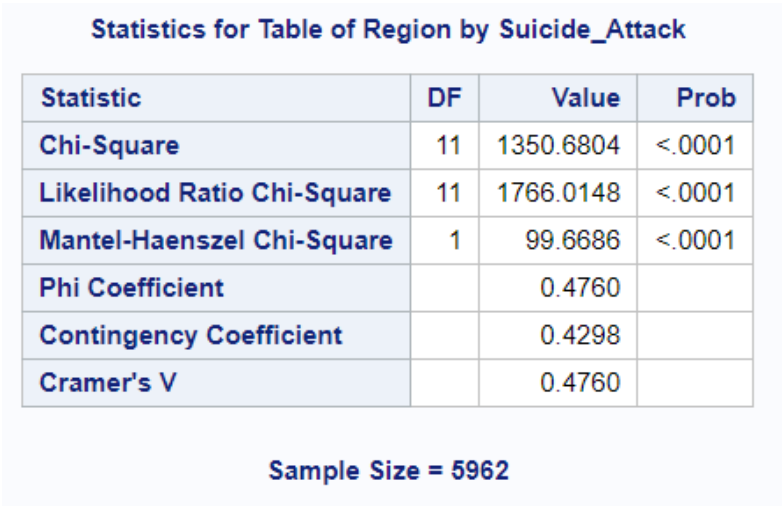


Figure 2.2



Statistics for Table of Region by Suicide_Attack

Statistic	DF	Value	Prob
Chi-Square	11	1350.6804	<.0001
Likelihood Ratio Chi-Square	11	1766.0148	<.0001
Mantel-Haenszel Chi-Square	1	99.6686	<.0001
Phi Coefficient		0.4760	
Contingency Coefficient		0.4298	
Cramer's V		0.4760	

Sample Size = 5962

Figure 2.3

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1054.36329	175.72722	26.06	<.0001
Error	5955	40160	6.74383		
Corrected Total	5961	41214			

Root MSE	2.59689
Dependent Mean	0.64458
R-Square	0.0256
Adj R-Sq	0.0246
AIC	17350
AICC	17350
SBC	11433

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.050465	0.332666	0.15	0.8794
Number_of_Terrorists	1	0.003283	0.000553	5.94	<.0001
Weapon_Type1 Chemical	1	-0.128642	0.873527	-0.15	0.8829
Weapon_Type1 Explosives	1	0.625381	0.125694	4.98	<.0001
Weapon_Type1 Firearms	1	-0.140641	0.133494	-1.05	0.2921
Weapon_Type1 Incendiary	1	-0.186205	0.156359	-1.19	0.2337
Weapon_Type1 Melee	0	0	.	.	.
Extended_24hrs 0	1	0.294197	0.318579	0.92	0.3558
Extended_24hrs 1	0	0	.	.	.

Figure 3.1

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4857.286	4807.990
SC	4863.979	4841.456
-2 Log L	4855.286	4797.990

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	57.2958	4	<.0001
Score	36.3360	4	<.0001
Wald	42.7902	4	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Number_of_Terrorists	1	11.3705	0.0007
Multiple_Attacks	1	7.2699	0.0070
Suicide_Attack	1	6.1238	0.0133
Gun_Used	1	15.3064	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	2.6879	0.1833	214.9863	<.0001
Number_of_Terrorists		1	0.0146	0.00434	11.3705	0.0007
Multiple_Attacks	0	1	-0.3295	0.1222	7.2699	0.0070
Multiple_Attacks	1	0	0	.	.	.
Suicide_Attack	0	1	-0.2121	0.0857	6.1238	0.0133
Suicide_Attack	1	0	0	.	.	.
Gun_Used	0	1	-0.5610	0.1434	15.3064	<.0001
Gun_Used	1	0	0	.	.	.

Figure 4.1

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	456.44297	65.20614	23.11	<.0001
Error	5954	16801	2.82187		
Corrected Total	5961	17258			

Root MSE	1.67984
Dependent Mean	0.46662
R-Square	0.0264
Adj R-Sq	0.0253
AIC	12157
AICC	12157
SBC	6246.51773

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.526795	0.238524	6.40	<.0001
Multiple_Attacks 0	1	-0.238402	0.064047	-3.72	0.0002
Multiple_Attacks 1	0	0	.	.	.
Attack_Type Armed Assault	1	-0.916953	0.234324	-3.91	<.0001
Attack_Type Assassination	1	-0.873192	0.241854	-3.61	0.0003
Attack_Type Bombing/Explosion	1	-0.997700	0.233027	-4.28	<.0001
Attack_Type Facility/Infrastructure Attack	1	-0.151067	0.241165	-0.63	0.5311
Attack_Type Hijacking	1	-0.610557	0.369126	-1.65	0.0982
Attack_Type Hostage Taking	1	-0.521079	0.261883	-1.99	0.0467
Attack_Type Unarmed Assault	0	0	.	.	.

Figure 5.1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4810.118	4536.825
SC	4816.811	4717.540
-2 Log L	4808.118	4482.825

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	325.2928	26	<.0001
Score	307.4056	26	<.0001
Wald	250.9325	26	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Attack_Type	6	161.9783	<.0001
Target_Type	20	73.1017	<.0001

Figure 6.1

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.7682	0.7762	12.7188	0.0004
Attack_Type	Armed Assault	1	0.3629	0.6047	0.3601	0.5484
Attack_Type	Assassination	1	-0.9038	0.6718	1.8099	0.1785
Attack_Type	Bombing/Explosion	1	1.4508	0.5997	5.8529	0.0156
Attack_Type	Facility/Infrastructure Attack	1	1.2741	0.6092	4.3746	0.0365
Attack_Type	Hijacking	1	1.3008	0.8354	2.4246	0.1194
Attack_Type	Hostage Taking	1	0.4643	0.6524	0.5065	0.4767
Attack_Type	Unarmed Assault	0	0	.	.	.
Target_Type	Abortion Related	1	-0.00082	0.5687	0.0000	0.9988
Target_Type	Airports & Aircraft	1	-1.0316	0.8135	1.6081	0.2048
Target_Type	Business	1	-0.00922	0.5051	0.0003	0.9854
Target_Type	Educational Institution	1	-0.0446	0.5454	0.0067	0.9348
Target_Type	Food or Water Supply	1	-11.6392	306.3	0.0014	0.9697
Target_Type	Government (Diplomatic)	1	-0.4936	0.5991	0.6788	0.4100
Target_Type	Government (General)	1	-0.3719	0.5106	0.5304	0.4664
Target_Type	Journalists & Media	1	-0.2964	0.6214	0.2276	0.6333
Target_Type	Maritime	1	-0.9841	1.1608	0.7187	0.3966
Target_Type	Military	1	-0.2548	0.5032	0.2564	0.6126
Target_Type	NGO	1	-0.3737	0.7900	0.2238	0.6362
Target_Type	Other	1	1.9497	1.5346	1.6143	0.2039
Target_Type	Police	1	-0.3763	0.5070	0.5509	0.4580
Target_Type	Private Citizens & Property	1	0.1936	0.5004	0.1496	0.6989
Target_Type	Religious Figures/Institutions	1	0.1425	0.5140	0.0769	0.7816
Target_Type	Telecommunication	1	1.5460	0.5786	7.1391	0.0075
Target_Type	Terrorists/Non-State Militia	1	0.1908	0.5466	0.1219	0.7270

Figure 6.2

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	65.6	Somers' D	0.364
Percent Discordant	29.2	Gamma	0.384
Percent Tied	5.1	Tau-a	0.087
Pairs	4255257	c	0.682

Figure 6.3

REFERENCES

https://www.kaggle.com/START-UMD/gtd

https://www.washingtonpost.com/news/monkey-cage/wp/2014/08/06/the-challenges-of-collecting-terrorism-data/?utm_term=.4427aa4c9c95

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5633603/