Model Results and Interpretation

Correlation Coefficients

#Correlation Coefficients				
> cor(conflict)	Fatalities	Events	Sources	Actors
Fatalities	1	0.629995	-0.1427	-0.13652
Events	0.629995	1	0.077298	-0.00676
Sources	-0.1427	0.077298	1	0.639146
Actors	-0.13652	-0.00676	0.639146	1

The table above shows the correlation coefficients between the Y and X variables in the conflict dataset. The Y variable here is Fatalities. Events, Sources, and Actors are X variables. As contained in the table:

- There is expectedly a perfect positive correlation between fatalities and fatalities, events and events, sources and sources, and actors and actors.
- There is a strong inverse relationship between sources and fatalities, and actors and fatalities. Ideally, there may be no point continuing with the insignificant variables.
- However, the correlation coefficient of .62 between events and fatalities show a strong positive relationship between the two variables. This suggests that as the number of events increase, the number of fatalities also increase by 62 percent.

Before testing the regression against the entire table, I checked the regression for fatalities and events. fatalities=(slope)event+(intercept)

Coefficients:				
	Estimate	Std. Error	t Value	Pr(>|t|)
(Intercept)	26.3206	61.1303	0.431	0.669
events	0.9007	0.1756	5.131	7.80E-06 ***
Multiple R-squared: 0.3969		Adjusted R-squared: 0.3818		
F-statistic: 26.32 on 1 and 40 DF, p-value: 7.80E-06				

With a multiple r-squared of .3969, the table above shows events <0.05 p-value at 7.80E-06 *** between events and fatalities. This means that the events model explains 39% of the variability of fatalities and has a significant contribution to fatalities.

Notwithstanding, it is best practice to proceed to regression analysis when dealing with a multiple linear regression. Hence, the regression statistics below.

Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	98.6947	77.0135	1.28E+00	0.208
Events	0.9207	0.1755	5.246	6.13E-06 ***
Sources	-1.9864	1.7362	-1.144	0.26
Actors	-0.234	2.4089	-0.097	0.923
Residual standard error: 606.4 on 43 degrees of freedom				
Multiple R-squared: 0.4339		Adjusted R-squared: 0.3892		
F-statistic: 9.708 on 3 and 38 DF, p-value: 6.88E-05				

Since this is multiple linear regression, I check for adjusted R-Squared to determine the strength of my model, determine best fit.

- Based off of the above table, Events, with a p-value of 6.13E-06 ***, has a more significant relationship with fatalities. The strength of this model is further checked by an Adjusted R-squared value of 0.3892 and a 6.88E-05 p-value. However, I proceeded to check for Archaic Information Criteria (AIC).
- I used ‘step’ to determine the AIC in order to check multicollinear variables and remove the insignificant variables and their effects.
- From the analysis, AIC is lowest at 484.36 when regression is between Fatalities (y) and Events (x1) and Sources (x2).
- To determine the strength of this model, the summary of AIC model revealed an Adjusted R-Squared of 0.4047 and a p-value of 1.526e-05, with Events having a p-value of 4.32e-06 *** as against a p-value 0.119 in the relationship with Sources.

Hence, contrary to my PowerBI submissions that fatalities are influenced by the number of actors, this analysis shows that Events strongly explain the variability in the number of fatalities.