Statistical Analysis of FEMA-recorded Disasters from 1953 to 2025

Lilith Carpenter

CSCI-3328

May 5th, 2025

The Federal Emergency Management Agency (FEMA) is home to a plethora of records about each individual federally declared natural disaster. It provides records of each incident starting from 1953 all the way up to the current year of 2025. The data is separated by state (or territory), incident type, and date, allowing each incident to be bucketed into the year it occurred, the state it occurred in, and what type of disaster it was. This report aims to use various statistical processes to analyze the data and further create inferences based on it.

Currently, there are 64 total US states and territories, however only 59 of them have had a previously recorded disaster. For observing any arbitrary category of disaster, this gives us a sample space of 59, the total different locations where a disaster can occur. Therefore, if choosing from the set of all previously recorded disasters and assuming that each of the sample points are equally likely, the probability of observing a disaster that occurred in the Midwest would be 12/59. According to FEMA, there are 26 different events categorized as federally declarable disasters, meaning the total sample space of different disasters paired with different states would be 1,534.

Surface level analysis of the data yielded the average (μ), median (M), variance (σ^2), and standard deviation (σ) below.

$$\mu = 930.44 \text{ disasters}$$

$$M = 558 \text{ disasters}$$

$$\sigma^2 = 1{,}702{,}883.72 \text{ disasters}^2$$

$$\sigma = 1{,}304.94 \text{ disasters}$$

This data illustrates that the typical number of disasters per year is around 930, but the extremely high variance shows there are large ranges for the data. This may be explained by the fact that the number of disasters declared per year has increased drastically since the mid-20th century when recording began.

When instead looking only at the past 10 years of recorded FEMA data, average (μ), median (M), variance (σ^2), and standard deviation (σ) was calculated again.

$$\mu = 2{,}273.7 \text{ disasters}$$

$$M = 1{,}586.5 \text{ disasters}$$

$$\sigma^2 = 6{,}685{,}697.79 \text{ disasters}^2$$

$$\sigma = 2{,}585.67 \text{ disasters}$$

The average number of disasters jumps to around 2274, with an even larger variance than entire data set. This trend in the data suggests many more disasters occurring within the US and its territories, or at least more disasters being large enough to require being recorded and federal aid from organizations like FEMA. The large variance is also consistent with the large outlier present in the data set. The year 2020 saw an explosion in biological incidents being declared as a result of the Coronavirus outbreak, inflating 2020's disaster rate to 9,490.

When first organizing the data, some other meaningful trends were noticed. The most common recorded incidents by far were severe storms, with hurricanes and floods following closely behind. Severe storms made up 28% of all recorded disasters, with hurricanes making up 20% and floods reaching 16%. The number of recorded incidents by year has been steadily trending upwards, but the years 2005 and 2020 stand out in the data because of the huge uptick in incidents from hurricane Katrina and the Coronavirus outbreak respectively. When going by state, Texas had the highest frequency of disasters, with almost double the number of disasters in Kentucky, the second most frequent location of recorded disasters. Texas disasters would make up around 8% of the total data, with Kentucky disasters at 5% and Florida at 4%.

From the previously mentioned statistic, 4% of all recorded disasters occurred in Florida. Furthermore, 2.1% of all the disasters were hurricanes in Florida. Using this data, it is possible to determine the likelihood that a disaster was a hurricane, given only the information that the disaster was observed in Florida. Using the formula for conditional probability,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

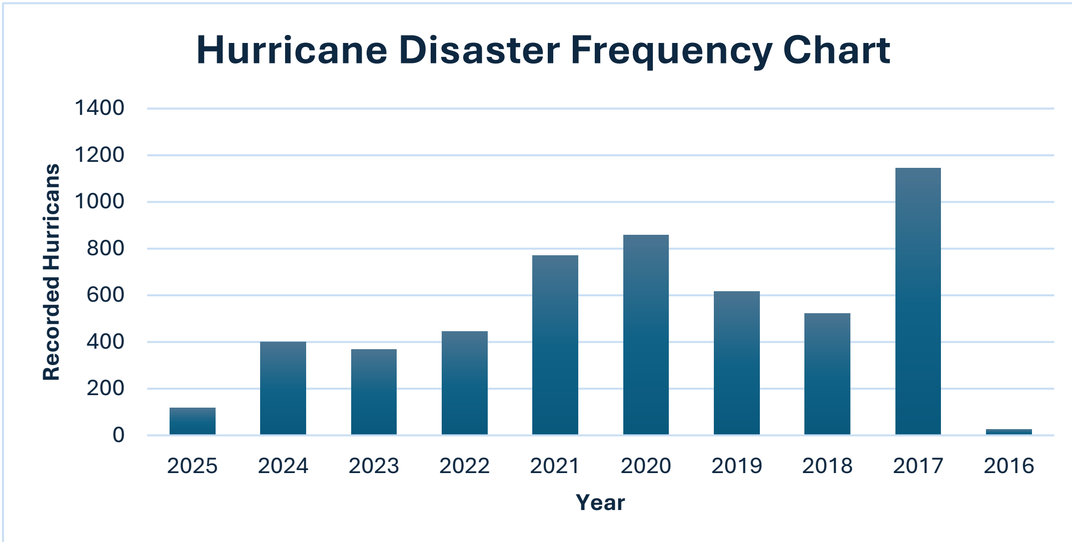
where event A describes observing a hurricane and event B describes observing any natural disaster in Florida. The result yields 52.5%, meaning that just by knowing that a disaster occurred in Florida, the probability of the disaster being a hurricane jumps to about 1 in 2.

In contrast to the previous statistic, the probability of a FEMA-declared hurricane occurring, given that the disaster was in any location other than Florida, rests only around 18.84%. Using this statistic in conjunction with the fact that 20% of all recorded disasters were hurricanes, it is possible to find the opposite probability, the probability that a disaster was in Florida given it was a hurricane. Using Baye's Theorem,

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

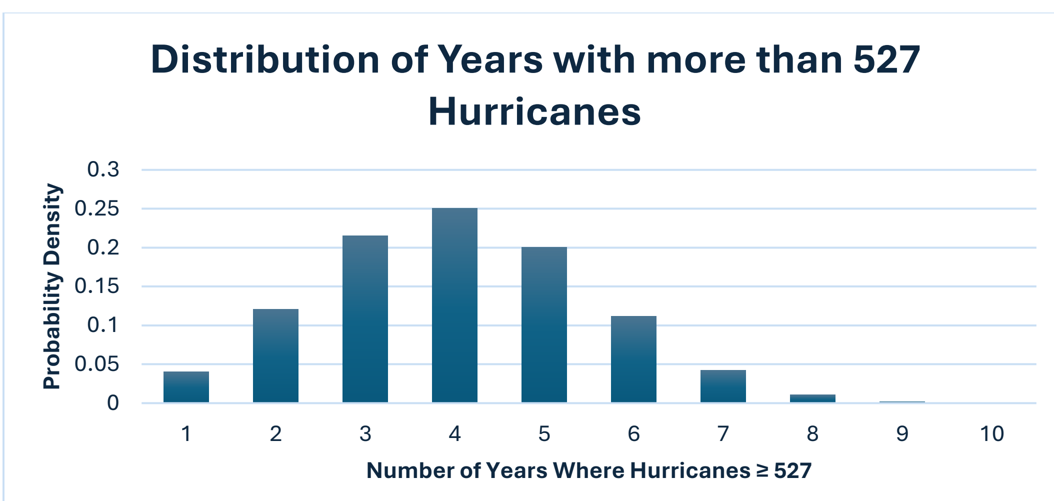
where event A and B describe the same events mentioned previously, the probability is found to be 10.6%. This means that just by knowing that a FEMA-declared hurricane occurred, the probability of that disaster occurring in Florida becomes about 1 in 10. These statistics are clearly able to show the sheer prevalence and routineness of hurricanes within the state of Florida.

Looking into FEMA-declared hurricanes recorded in the past 10 years, the histogram sorted by year is shown below.



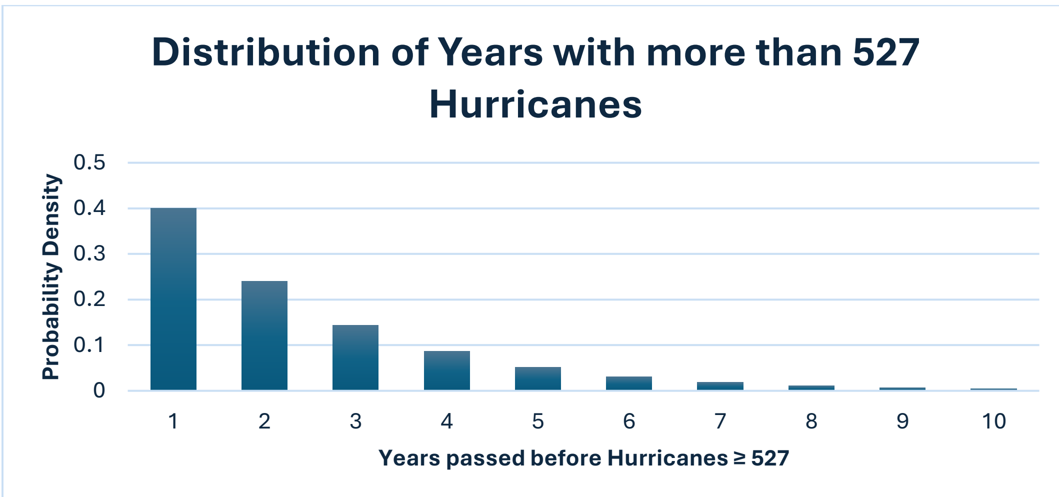
This data yields $\mu = 527$ for the amount of declared hurricanes in the US per given year. Using these past 10 years' probability distribution, it was made possible to calculate the probability of observing a year that has less than average recorded hurricanes. Counting the number of observed years with less than 527 hurricanes, constituting 6 years total, meaning a 60% probability of observing a year with less hurricanes than average.

Then, using this statistic, it is possible to use the binomial distribution to make some predictions about the next 10 years' recorded hurricane numbers. With the probability of observing a single year with more than average hurricanes being 40%, the following probability distribution was recorded.



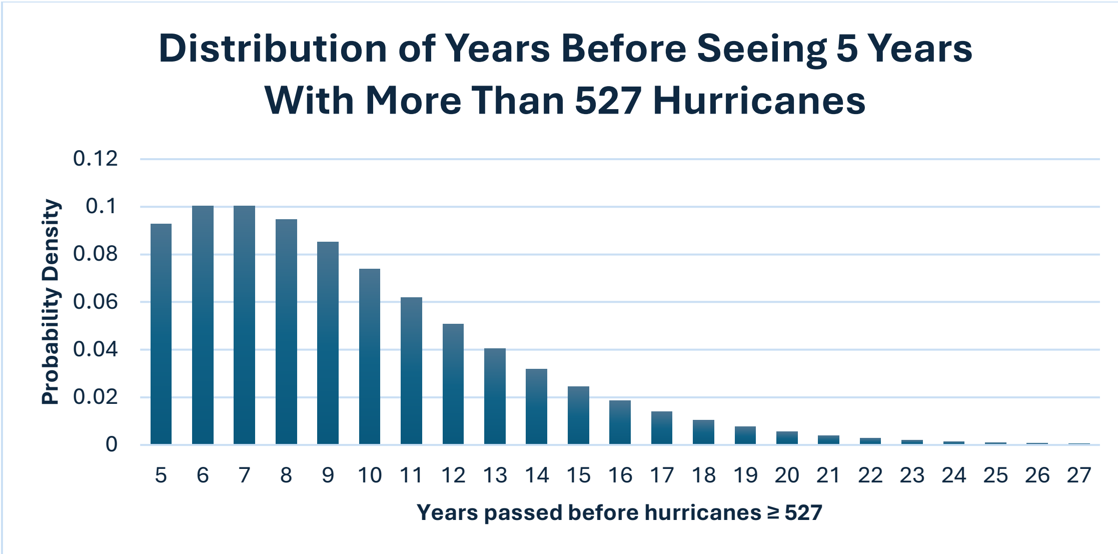
This graphic visualizes the individual probabilities for the various amounts of years recorded with more than average hurricanes in the next 10 years. The data follows the binomial distribution, under the assumption that the previous year's number of hurricanes does not affect the next, where the most likely number of years to be recorded with more than average hurricanes is 4 years out of 10. This result means that the probability of observing just 1 year with more than average hurricanes and 7 years with more than average hurricanes is extremely similar, with a probability of only about 4%.

Using the same previous statistic, it is then possible to use the geometric distribution to determine the individual probabilities for how many years it will take for more than 527 hurricanes to be recorded.



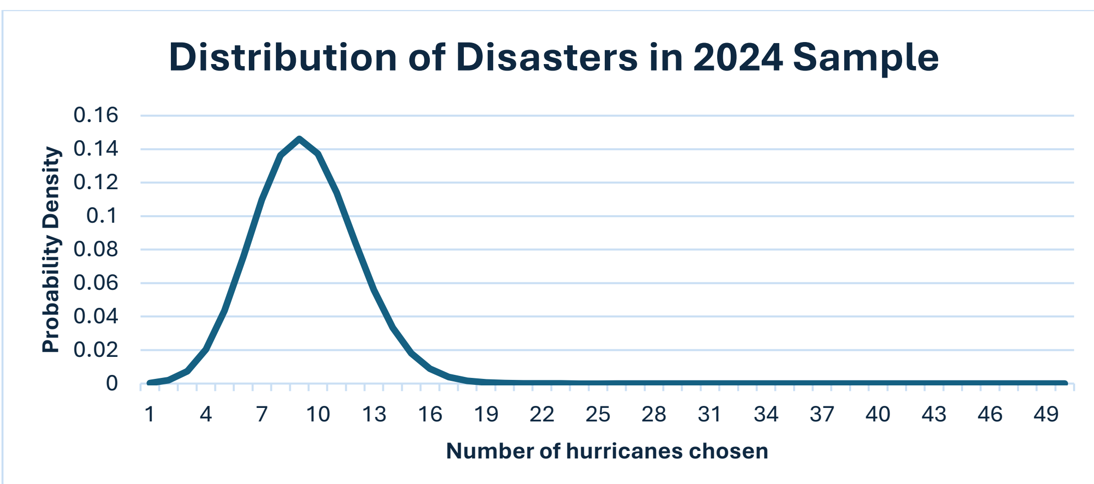
This graphic describes the individual probabilities of the number of years passed before seeing a year with more than average hurricanes. It shows that the probability of going 1 year before seeing above average hurricanes is about 40%, then going 2 years is about 25%, getting smaller as the years increase. The distribution is able to demonstrate that even though the probability of a year being recorded with above average hurricanes is less probable than below average, the probability of going long stretches of years without encountering above average hurricane numbers is improbable.

Still using the previous statistic, it is also possible to use another distribution, assuming the data follows the negative binomial distribution, to determine specific probabilities of the number of years to pass before seeing a desired number of years with above average hurricanes declared. To observe how many years it would take before seeing 5 years with above average hurricane records, the below probability distribution was recorded.



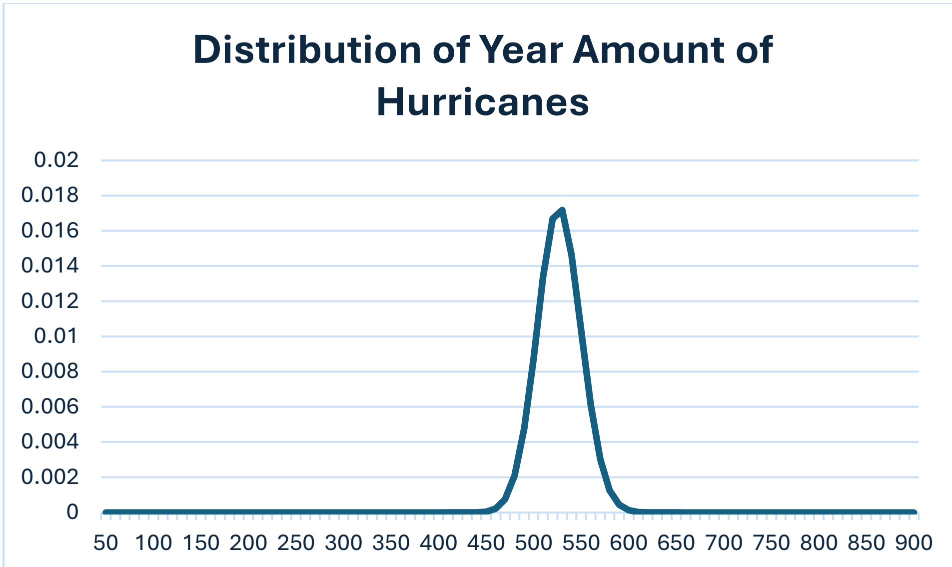
In conjunction with the graph, it is easy to visualize how many years are expected to pass before seeing 5 years with above average recorded hurricanes. $\mu = 7.5$, Meaning about 7 years are expected to pass before seeing 5 years with above average hurricanes. This statistic makes demonstration of the fact that even though it is more likely to see a year with below average hurricanes, it would only take about 7 years to see 5 years with above average hurricanes.

FEMA data from the year 2024 indicates 400 recorded hurricanes out of 2147 total disasters. The hypergeometric distribution allows for choosing a random disaster without replacement. When choosing 50 of the 2024 disasters at random without replacement, it is possible to calculate the probability of choosing a number of hurricanes out of a total sample.



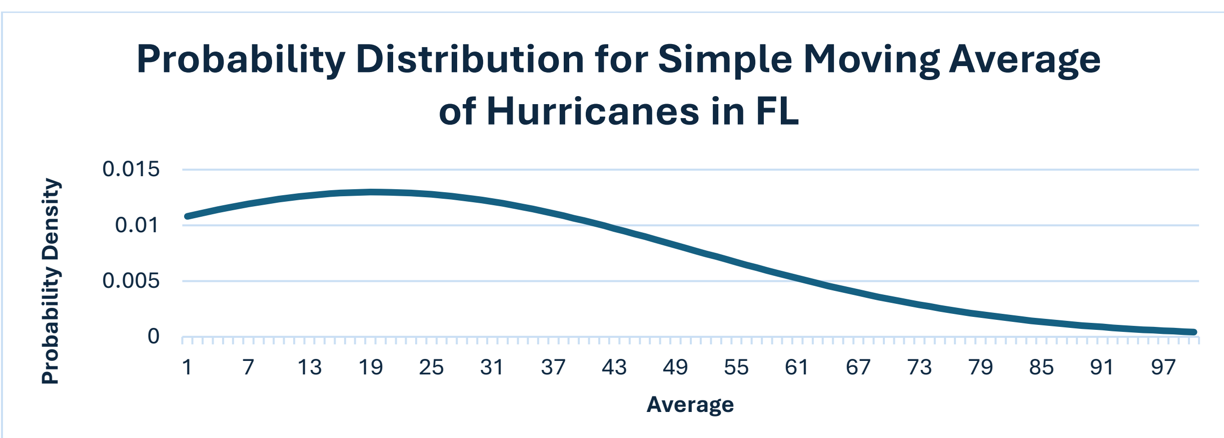
This distribution yields $\mu = 9.3153$, meaning the average number of hurricanes chosen out of 50 for the year 2024 would be about 9.

Then, using the Poisson distribution in conjunction with the expected recorded hurricane amount, it is then possible to calculate the probability distribution to see the exact amounts of hurricanes occurring, assuming that this data follows the Poisson distribution.



Inferences from this data can show the probability that 3 years in a row have no more than 530 recorded hurricanes in a year, where the probability of a single year having 530 hurricanes or less is 56.34%. This can be found using the mn rule, multiplying this probability by itself twice. This results in 17.88%, the likelihood that 3 years will go by with no more than 530 hurricanes.

A simply moving average of hurricane amounts in Florida was created using 5 years as the window. Assuming that the data follows the normal distribution, the following graph was created.



Using the graph above, the result of $\mu = 19.64$ and $\sigma = 30.73$ are obtained. It is also made possible to define the mean amount of hurricanes of any given 5-year interval as a continuous random variable. Given the previous assumption that the data follows the normal distribution, it can be defined with a probability density function as follows,

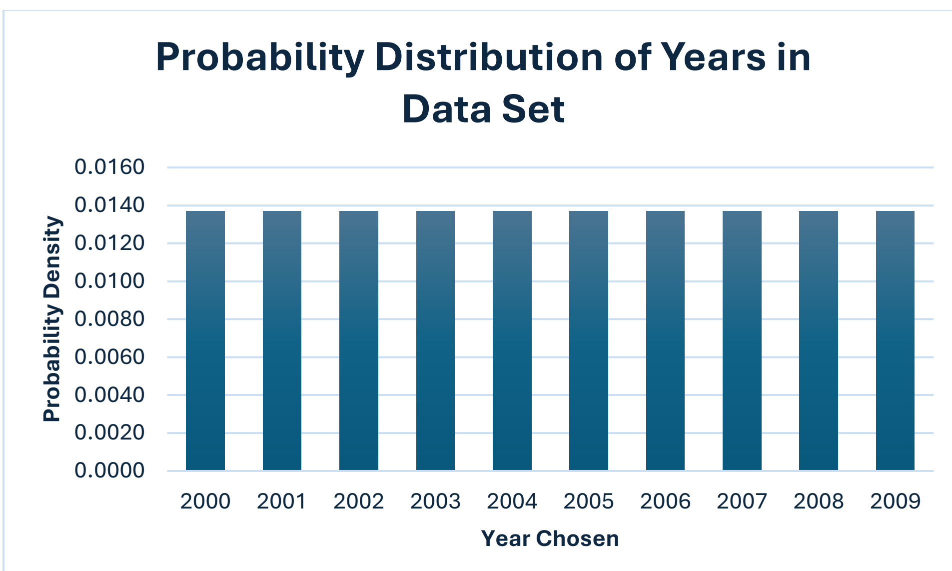
$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}$$

With this probability density function, it allows for the probability that any given 5-year interval has less than the average number of hurricanes to be obtained.

$$P(Y < 19.64) = \int_0^{19.64} \frac{1}{30.73\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{19.64-19.64}{30.73})^2} dy = .2549 = 25.29\%$$

Essentially, this statistic aims to say that for any given 5-year interval in Florida, the probability that the μ obtained for those 5 years is less than the grand mean of 19.64 is 25.29%.

The number of years included in the data sets follows the uniform distribution. Therefore, when choosing any given year at random, each year has an equal probability of being the year chosen. The uniform distribution allows for calculation of each individual probability given the fact that there are 73 years total.



The graphic above clearly illustrates each individual year's probability of being chosen in the set of the 73 total recorded years is exactly the same no matter the year.

Using joint probability functions, it is also possible to investigate the probability of two different events occurring. Approximately, the number of recorded hurricanes and fires in Texas in the past 10 years are nearly the same, leading to the assumption that they have similar probabilities of occurring. It is possible to calculate the number of disasters chosen being either fires or disasters when choosing 2 disasters at random.

$$F = \text{Fire disaster was chosen}$$

$$H = \text{Hurricane disaster was chosen}$$

$$N = \text{Neither disaster was chosen}$$

$$S = \{F,F\},\{F,H\},\{H,F\},\{H,H\},\{F,N\}.\{N,F\},\{H,N\},\{N,H\},\{N,N\}$$

Let X be the number of hurricane disasters chosen and Y be the number of fire disasters chosen. With the sample space provided above, the following chart was created.

			X	
		0	**1**	**2**
	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$
Y	*1*	$\frac{2}{9}$	$\frac{2}{9}$	0
	2	$\frac{1}{9}$	0	0

Using the table above it is then possible to obtain the marginal probabilities for each of the variables. The only calculation to be made is adding the columns or rows for the variables.

X

$P(0)$	$P(1)$	$P(2)$
$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

Y

$P(0)$	$P(1)$	$P(2)$
$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

From these obtained marginal distributions, the dependency of X and Y can also be investigated. The formula for independent variables is as follows, $p(x, y) = p(x) \cdot p(y)$, which when substituted with values from the previous table shows the dependency of the variables, concluding that the variables are dependent on one another.

$$p(1,1) = p_x(1) \cdot p_y(1)$$

$$\frac{2}{9} = \frac{4}{9} \cdot \frac{4}{9}$$

$$\frac{2}{9} \neq \frac{16}{81}$$

Much of the statistics found within this analysis seemed to be expected based on the frequency of where certain disasters appear and how often they occur. It is also important to note that the number of declared disasters of any type may not indicate individual events. If the same disaster, such as a hurricane, occurs in two different counties but comes from the same storm, the hurricane disaster will be recorded twice by FEMA's records. Each of these individual events should only be thought to represent an individual instance of a county requesting government assistance in the face of a natural disaster.

The only thing the statistics in the data found suggests is that government assistance provided to communities in the face of natural disasters is extremely important, seeing the frequency of how often they occur. As FEMA faces large funding cuts and even the possibility of complete shutdown in the current political climate, these issues raise concern for disaster preparedness and response in disaster heavy states like Texas and Florida. FEMA's disaster relief provides extreme support to US locations where disasters are frequent, and after extremely destructive hurricanes like Irma, organizations are proved necessary countlessly.