Lightning vs. Human-Ignited Fires: A Bayesian Comparison

Abigail Coelho
December 8, 2022
University of California, Los Angeles

Abstract

Fires are a prevalent aspect of living in California. This research examined the differences in the ignition conditions of fires caused by lightning compared to those caused by humans. A Bayesian method was implemented to compare the distributions of the variables of the two groups. The distributions of the ignition day's maximum temperature and average windspeed, and the location's distance from highway were predicted. The results demonstrated a significant difference between the two groups.

Introduction

Moving out of the hills to a friend's couch, watching from the other side of the highway, coming home to find that nothing remains of the neighborhood—these are ways in which various members of my family, not to mention the experiences of my friends, have experienced the increase in wildfires. Human activity leading to climate change has increased the number and intensity of wildfires, but beyond climate change, human activity can also directly lead to fires being ignited. A neglected campfire, tossed cigarette, or crashed drone can lead to thousands of acres burned, but there are also fires caused by natural lightning strikes. Do these fires share the same conditions for igniting a fire? Should smokey the bear always set the risk of fires to high for humans no matter the weather?

This goal of this research was to determine if the conditions for human caused fires are different from those of naturally caused fires. The weather variables that were examined were the maximum temperature and average wind speed for the day the fire ignited. The third variable examined was the ignition location's distance to highway, a proxy for an increase in human activity, providing more opportunities for humans to cause fires. To test the hypotheses, a Bayesian method was implemented based on the method from "Bayesian Estimation Supersedes the t-Test" (Kruschke).

Data

The fire occurrence dataset was collected from the USDA Forest Service website for three years from 2010 to 2012 in California. The dataset includes all known fires of any size ignited on National Forest System Lands, and the variables include the ignition date, location, and cause. This dataset was combined with daily weather station data, containing the station location, daily maximum temperature and windspeed from the NOAA. Each fire was connected to the closet weather station, and the weather data for the ignition day was matched to each fire. Lastly, geospatial data on the locations of highways was gathered from USDOT, and the closest distance to a highway from the ignition location was computed.

The dataset was divided into two groups using the fire cause variable. The first group consisted of fires ignited by lightning and contained 1,280 samples. The second group, fires ignited by humans, contained 2,654 samples. The large sample size allowed for a high precision of parameter estimate, due to sample noise decreasing.

Bayesian Estimation Method

Three hypotheses were tested. Each hypotheses examined whether there is a difference in the distributions of the two groups—naturally or human-ignited fires—for each of the three variables—maximum temperature, average windspeed, and distance from highway. However, instead of the frequentist approach of relying on t-tests to reject a null hypothesis, a Bayesian method was implemented. Unlike the frequentist method, this Bayesian method provides more information regarding the distributions of the two groups beyond rejecting or failing to reject a null hypothesis. The three hypotheses were:

1. Fires ignited by lightning have a different *maximum temperature* distribution than fires caused by humans.
2. Fires ignited by lightning have a different *average windspeed* distribution than fires caused by humans.
3. Fires ignited by lightning have a different *distance from highway* distribution than fires caused by humans.

The method used in this research application came from "Bayesian Estimation Supersedes the t-Test" (Kruschke), and the paper's BEST package for R was used for implementation. The default assumption from the BEST method assumes that likelihood of the data follows a t-distribution with three parameters, the location, standard deviation, and normality. The t-distribution is advantageous over a normal distribution because the tails can have more weight, providing more room for outliers. A t-distribution was appropriate for the applications to temperature and windspeed. For distance, the likelihood function was modified to a log-normal distribution; being able to modify the assumed distributions is a powerful advantage of Bayesian methods.

The priors for the parameters of the t-distribution likelihood were set to uninformed in the BEST package. The prior of the mean of each group was assumed to be a normal distribution with a mean of the pooled data and a standard deviation of 1,000 times the pooled data. The prior for the standard deviation was a uniform from 1000 times to 1/1000 of the standard deviation of the pooled data. Lastly, the normality parameter had a gamma prior with mean and standard deviation of 30. The same priors were again used for the location and scale parameters of the log-normal distribution, excluding normality which is not a parameter of log-normals.

Another advantage of the Bayesian methods is that the priors of the parameters can be changed to informed priors, but uninformed priors are appropriate for this analysis because I have minimal knowledge about the conditions of when wildfires are ignited. Informed priors would be beneficial when provided by subject-matter experts, and the BEST method could be modified to use known trends.

The method from Kruschke's paper then uses MCMC sampling to generate a large number of samples of possible values from the marginal posterior for each parameter. These samples of the parameters from the posterior are then be plotted to visualize the likely values of the parameters. Kruschke recommends returning 100,000 samples from the Markov chain, but for efficiency the chain samples were reduced to 1,000 with 100 steps removed for burn-in and the chain thinned to every 10 steps to reduce autocorrelation.

Hypothesis 1: Maximum Temperature

The first hypothesis—that lightning-caused fires have a different distribution of the daily maximum temperature for those caused by humans—was found to be true by this Bayesian method. The 1,000 chain samples of the marginal posterior for each parameter of the two groups were plotted as histograms. The distribution of the marginal posterior for the mean of lightning fires did not seem to overlap with that of the human-caused fires (figure 1, "Group X Mean"). The histogram of the differences of the means did not contain any values below 2, and the lack of any values close to the region of practical equivalence (ROPE) shows that the group means are indeed significantly different (figure 1, "Difference of Means").

The differences in the credible standard deviations of the two groups also appeared to be significantly different (figure 1, "Group X Std. Dev." and "Difference of Std. Dev.s"). The normality marginal posterior values for both groups were centered around a log value of 1.1, which is close to a value of 10. This value indicated that likely both distributions should have smaller tails (figure 1, "Normality). Lastly, the effect size did have come values within the ROPE region, but most of the values are outside of the region, showing there is some relationship between the two group's values (figure 1, "Effect Size").

Samples of the posterior predictive distributions were plotted against histograms of the true values for both groups (figure 1, "Data Group X w. Post. Pred."). A posterior predictive check concluded that for lightning-caused fires, the predicted distribution fits well; however, for human-caused, the histogram of the data showed a skewed distribution which the t-distribution is unable to capture. The skew in the distribution may be because humans cannot comfortably be outside at temperatures above 100 but can wear appropriate clothing for cooler temperatures.

The fires ignited by lightning had a higher mean value and a smaller deviation than the human-caused fires. This may be due to the particular and somewhat-rare conditions for lightning to occur. Mother-nature does not follow this pattern. Additionally, the skew of the distribution of the human-caused fires may have resulted in the lower mean value. Comparing the means and medians of the groups shows that the median values are closer than the means (table 1).

Comparison of Sample Statistics		
	Sample Mean	**Sample Median**
Lightning	82.9	84
Human-caused	79.2	81

Table 1: The human-caused have a higher sample mean and median than lightning.

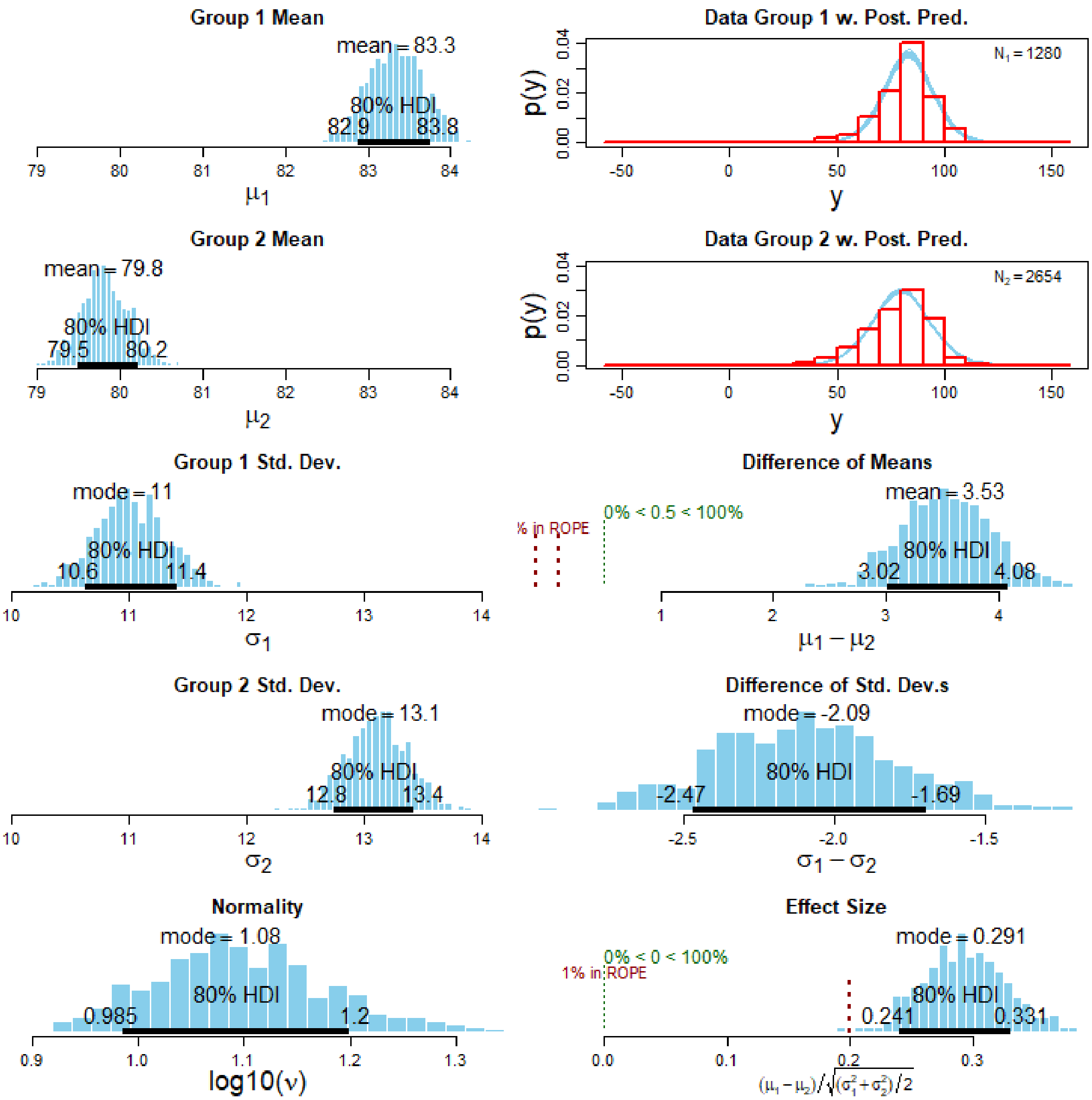

Figure 1: This figure is for the maximum temperature for lightning-caused (group 1) and human-caused (group 2). On the left, the histograms are plots of the 1,000 samples from the chain of the marginal posterior distribution for each parameter. The two top-right plots have a histogram of the true sample values overlayed on the predicted posterior distributions. The middle-right two graphs are of the difference in the groups mean and standard deviation, and the bottom-right is the effect size for the 1,000 chain samples.

Hypothesis 2: Average Windspeed

The results from the analysis on the average windspeed showed that the second hypothesis holds as well. The samples from the marginal posterior of the means for each group showed that the lightning fires have a smaller windspeed than the human caused (figure 2). The difference of the means did not contain any values in the ROPE area, so the values being significantly different can be accepted. When humans cause fires the wild may be slightly higher since wind causes fires to spread beyond barriers more easily. It would be interesting to compare the windspeed of when fires were ignited to general windspeeds from any day. The standard deviation predictions from the marginal posterior of the lightning-caused fires were again lower. Again, this may be do to the specific required conditions for lightning to occur.

Interestingly, for windspeed, the normality value of the t-distribution posterior had a smaller value than that of the maximum temperature, indicating that maximum temperature has a larger spread, requiring accounting for more outliers.

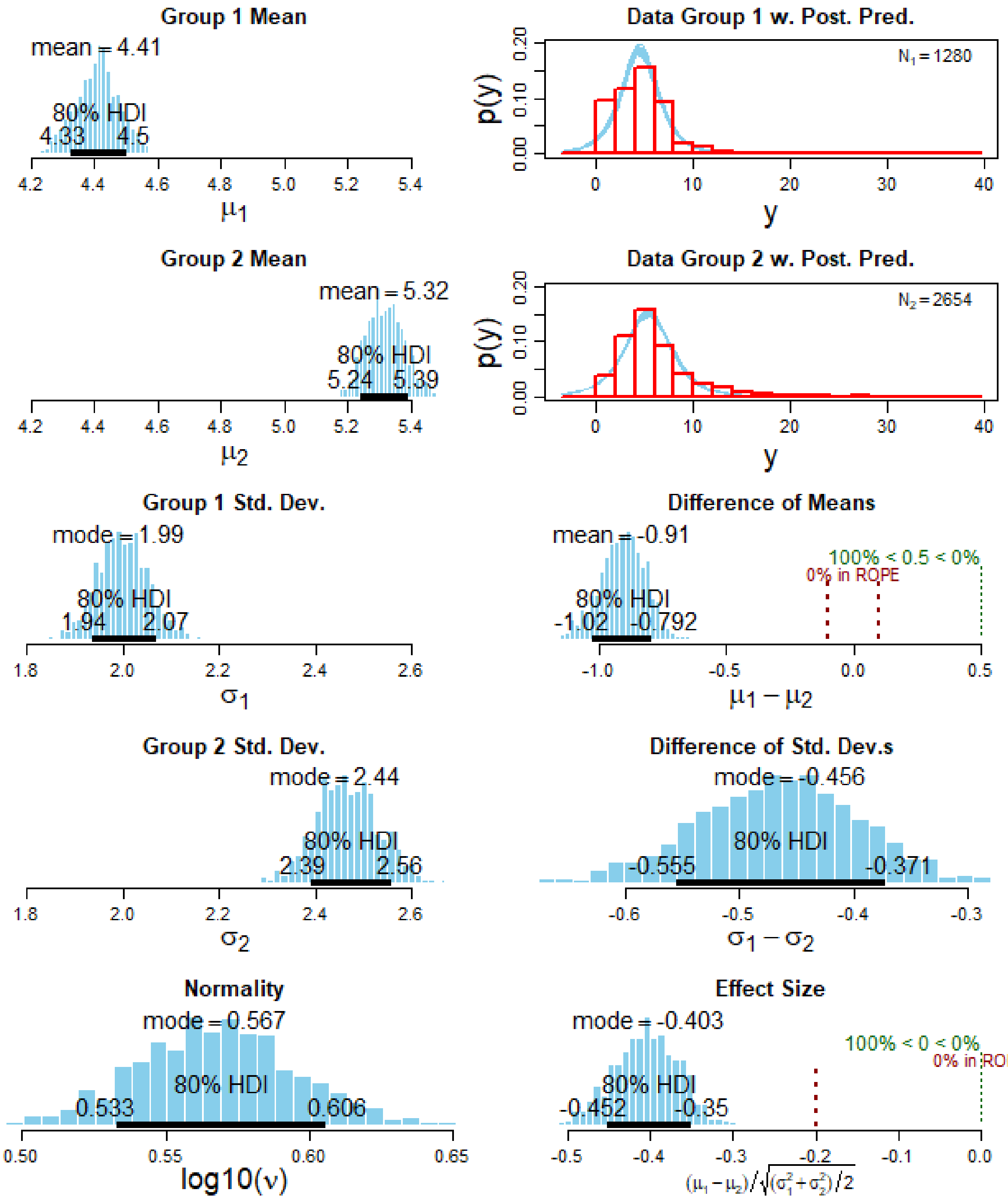

Figure 2: This figure is for the average windspeed for lightning-caused (group 1) and human-caused (group 2). On the left, the histograms are plots of the 1,000 samples from the chain of the marginal posterior distribution for each parameter. The two top-right plots have a histogram of the true sample values overlayed on the predicted posterior distributions. The middle-right two graphs are of the difference in the groups mean and standard deviation, and the bottom-right is the effect size for the 1,000 chain samples.

Hypothesis 3: Distance from Highway

The final hypothesis was that the distribution of the distance from the highway is different for fires ignited by lightning than humans. This hypothesis again was shown to hold true but required a modification of the prepackaged BEST functions. As previously discussed, instead of following a t-distribution as BEST assumes, the distances follow a distribution similar to a log-normal. Examining a density plot of the true sample values shows that—as expected—the values are skewed close to zero, and of course are non-negative (figure 3). As previously described, the BEST package was modified for a log-normal distribution with uninformed location and scale priors, and the same MCMC process was repeated for the modified program.

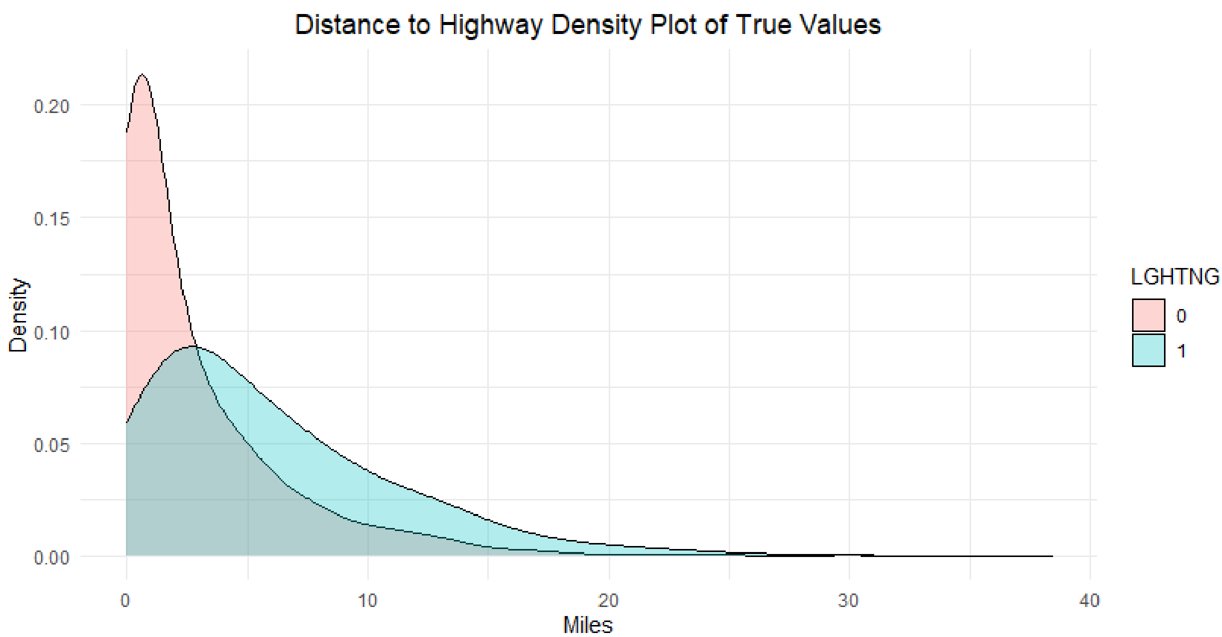

Figure 3: This plot shows the kernel densities of the true values of the distance from the highway for the two groups of fires.

The predicted posterior, using the mean value of the parameter chains, was examined (figure 4). The shapes of both groups of fires were very close to that of the density plot of the true sample values. However, the predicted distributions appeared to be too narrow compared to that of the samples. This may indicate that the prior for the scale parameter needed to be adjusted so that more weight can be given to distances over 5 miles.

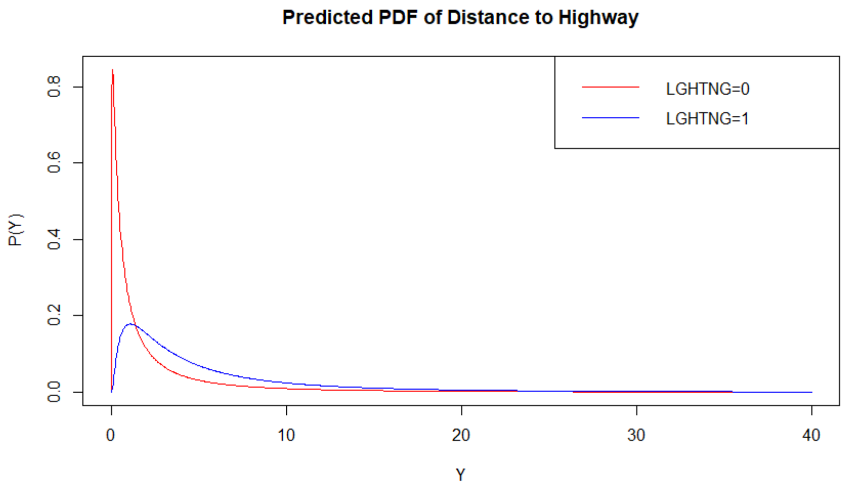

Figure 4: This is a plot of the predicted posterior, using the mean of the chain's parameter estimates, of the distance to the highway for fires.

The differences in the location and scale parameters were examined. The location parameter has higher credible values from the Markov chain for lightning fires compare to human-caused fires, and the difference in the location parameters, demonstrates that the difference likely does not ever touch a value of zero (figure 5). Human-cause fires are likely closer to roads because it is difficult for humans to get to locations away from roads.

The scale parameter is more challenging to interpret as the effects of real-world conditions, but the differences in the scale values were centered away from zero, indicating that there is a significant difference between the two distributions (figure 5).

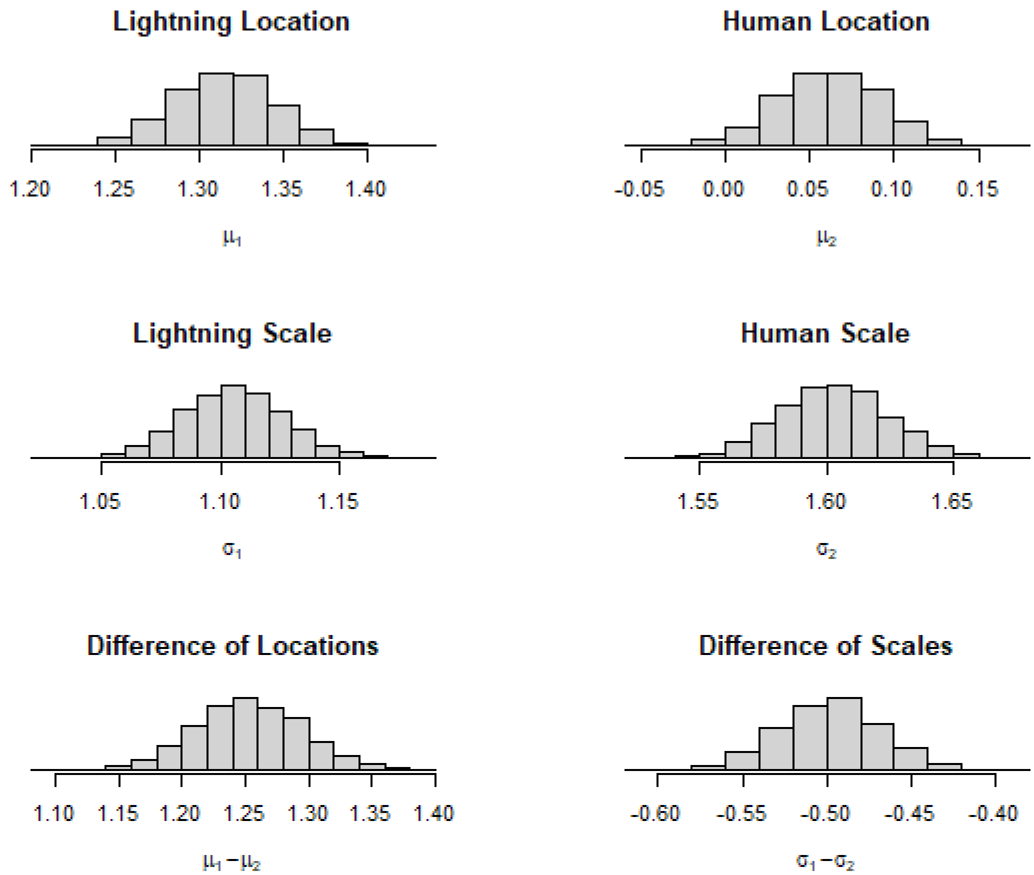

Figure 5: These are histograms of the chain samples for each marginal posteriors of the log-normal likelihood parameters for the distance from the highway for each group. The top two are for the location, followed by the scale, and the difference of the groups' parameters.

Discussion

All of the hypotheses were found to hold true; fires ignited by humans have different maximum temperature, windspeed, and distance from highway than those ignited by lightning. Human caused fires tend to occur over a larger range of maximum daily temperatures and have more chances of occurring when the windspeed is increased. Humans do not need the perfect conditions that lightening requires to cause a spark. This study further emphasizes the winder array of weather circumstances that allow humans to create the same environmental impact as compared to the narrow array for a natural phenomenon.

However, human-ignited fires do not stray far from where humans occur. Areas away from humas (far from roads) are much less likely to have fires caused by humans. This difference, combined with the differences in the weather variables, may indicate that for modeling the likelihood of fires occurring, there may be benefit in modeling fires caused by human and those cause by lightning separately.

Future Studies

If this research were to be repeated, testing different likelihood distributions would be beneficial. For maximum temperature, perhaps the normality variable should be different for both groups. Additionally, the distribution may be better fit to a distribution that does not require symmetry such as the log-normal. For the scale parameter for the log-normal, the uninformed prior should be updated to better capture a possible higher scale of values.

This research provides additional exploratory analysis for my thesis on special-temporal modeling of the likelihood of wildfires. When separately modeling the different causes of fires, this research has shown that the three variables discussed should have different values, depending on the cause.

Conclusion

The comparison of the distributions for maximum temperature, average windspeed, and distance from the highway for lightning versus human-ignited fires showed that the differences are significant. Furthermore, the Bayesian method better quantified the differences beyond concluding to reject a null hypothesis.

Citations

Bureau of Transportation Statistics. “North American Roads.” U.S. Department of Transportation. 14 April 2022. https://data-usdot.opendata.arcgis.com/datasets/usdot::north-american-roads/explore?location=34.414877%2C-109.495150%2C3.73

Forest Service. “Fire Occurrence FIRESTAT yearly.” US Department of Agriculture. 14 April 2022. https://data.fs.usda.gov/geodata/edw/edw_resources/meta/S_USA.Fire_Occurrence_FIRESTAT_YRLY.xml

Kruschke, John K. "Bayesian estimation supersedes the t test." Journal of Experimental Psychology: General 142.2 (2013): 573.

National Centers for Environmental Information. “Daily Summaries.” National Oceanic and Atmospheric Administration. 29 April 2022. https://www.ncdc.noaa.gov/cdo-web/search