Tri Cao

Dr. Jones

DATA 363

9 December 2020

**The Most Dangerous Place to Live During a Meteor Shower**

**Introduction**

Meteors have played a crucial role throughout history from Greek and Roman Mythology to romantic date nights in the countryside. However, meteors can have devastating effects if they are large enough and land in a densely populated area. In this project we will discuss the world’s most common meteor hotspots based on the average mass of meteorites that land within a given area to predict the most dangerous place to live. The data that we have investigated is from a free public NASA dataset that contains Meteoroid data.

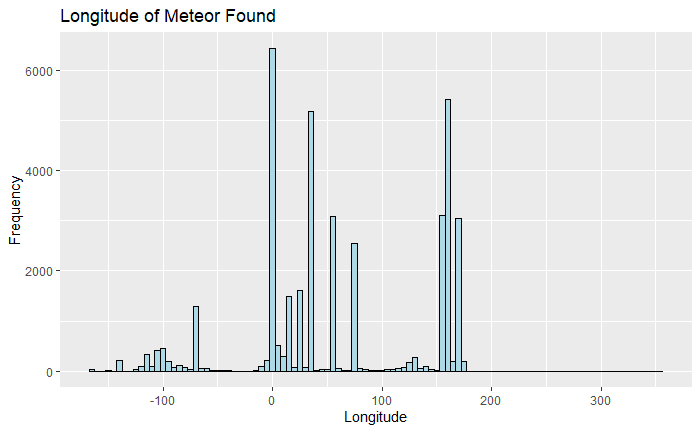
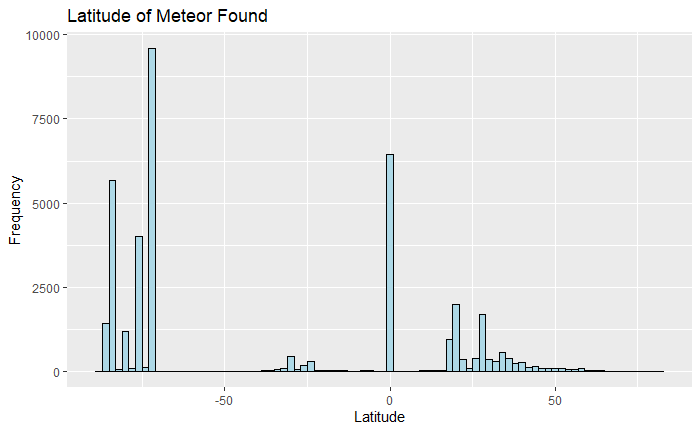
Many people are unaware that there is a difference between the types of Meteoroids. A Meteoroid is a piece of space rock ranging from grains of dust to small asteroids, but there are three different stages to each space rock: Meteoroid, Meteor, and Meteorite. Meteoroids are classified as Meteors when they enter the Earth’s atmosphere and burn up, so every shooting star you may see in the night sky is a meteoroid. Space rocks are called Meteorites when they impact the ground. Throughout this dataset we will be focusing on the meteorites since they can reach the ground and pose the greatest threat to those living on the surface of Earth. Throughout this project we want to see if there is any correlation between the mass of the meteorite and the location that it was found. If there is a location or continent that tends to get more harmful meteorite action it can be helpful for people in that area to know so they can either get insurance or move to a different location that would be safer from space rock activity.

**Methods**

The data that we used was collected by NASA and The Meteoritical Society. We found this data set online on a public database. This dataset was used because it seemed interesting and there was a lot of information within this dataset. The dataset has information about each meteorite containing the name, class of the meteorite, the mass in grams, the fall type, date, and geographical location. For this project we will primarily be using the geolocation and mass from the dataset and divide them into location categories. We would then count the frequency of landings per each categorical variable and find the average mass and the standard deviation of the mass that fall into these categories.

The dataset's reliability can be ensured because it was released by The Meteoritical Society and then reviewed and approved for public use in the United States by NASA. The Meteoritical Society is full of researchers from across the planet who specialize in planetary sciences. When the researchers discover a meteorite, they note the location and perform all the necessary measurements. The data these researchers collect is accurate as they do whatever they can to uphold the integrity of the meteorite measurements and the society’s reputation. There is some bias in this dataset, the researchers are limited only to locations that they can reach and stay long enough to perform analysis. This means that the dataset may not have as many points in the middle of the ocean and remote regions of the world. This would skew the data because we may not know if there are outstanding statistics in these locations. This form of bias is considered Sample bias. The actual data is not fully reflected in the dataset. Considering meteorologists can only measure the mass and location of meteorites they have found; any data of unknown meteorites would not be included in the dataset.

**Results**

We are seeking to explore if there is an observed association or any statistical relationship between the two variables: location and mass. The visualizations that I have currently are the frequency that a meteorite was found at a given longitude and latitude. These visualizations show that there are some locations that have much more meteorite activity than other locations. This could be due to the size of the meteorite because when there are a lot in one location the rock tends to be dust-like measuring in fractions of grams. Since the frequency does not directly correspond to dangerous living conditions the analysis should be on the mass of the meteorites for each continent. This is determined because the more massive meteorite yields a much greater impact force which would cause more damage.

**Analysis**

To see if the mass of these meteorites has any correlation with their location, we set up an Analysis of Variance test (ANOVA Test). The ANOVA test will determine if the mean meteorite mass for one continent differs from the other continents. The hypotheses that we will be testing are below.

, where

For the test to work properly the data has been separated into continents. The location of the meteorite landings was plotted with their corresponding geo-coordinates and using R we can figure out the continent from these coordinates. The seven continents the data were grouped into are Africa, Antarctica, Asia, Australia, Europe, North America, and South America. A table with the frequency of meteorites at each continent is shown below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Africa | Antarctica | Asia | Australia | Europe | North America | South America |
| **Frequency** | 2837 | 22088 | 3525 | 644 | 578 | 1708 | 638 |

From this table we can see that there is an abnormally large number of meteorites found in Antarctica, however this is possibly due to having a large land mass and the frequency of the meteorite strikes at each location is not our primary concern since we want to see where the largest meteorites strike. With the frequency of the meteorites given we can now set up the table for our ANOVA test.

|  |  |  |  |
| --- | --- | --- | --- |
| **Continent** | **Frequency** | **Mean Mass (kg)** | **St. Deviation (kg)** |
| **Africa** | 2837 | 41.54 | 1265.5 |
| **Antarctica** | 22088 | 0.176 | 3.1809 |
| **Asia** | 3525 | 20.496 | 513.0955 |
| **Australia** | 644 | 80.24 | 1023.84 |
| **Europe** | 578 | 77.38 | 963.37 |
| **North America** | 1708 | 53.94 | 832.07 |
| **South America** | 638 | 228.76 | 2417.97 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variation** | **Degrees of freedom** | **Sum of squares** | **Mean of Squares** |
| **Between Groups** | 6 |  |  |
| **Residual** | 32011 |  |  |
| **Total** | 32017 |  |  |

= 65.23

Our F test statistic is 65.23. The p-value of this test statistic is computed using the R command: pf(65.23, 6, 32011, lower.tail = FALSE) which yields a p-value of . Since this value is incredibly small, we fail to reject the null at any level over . In the context of this hypothesis test, at least one of the mean meteorite masses for each continent in this dataset is statistically different than the others.

Upon preliminary investigation, looking at the data reveals that Antarctica has a much lower mean mass than the grand mean and South America's mean mass is much higher. For this project we want to see the most dangerous place to live so we will pick South America to analyze further because the mean mass is larger. The additional analysis requires contrast testing which will tell us if South America's mean meteorite mass is different than the means for every other continent. The contrast is found using the hypotheses:

(

Where

The test statistic is computed by taking the expected value of the hypothesis and dividing it by the product of and the square root of the variance. After inputting all the variables, the t-statistic for South America comes out to with a p-value of . Since this p-value is exceedingly small we reject the null hypothesis. This implies that the mean meteorite mass in South America is significantly larger than the means in every other continent.

**Conclusion**

To answer the preliminary question; the most dangerous place to live if a meteor were to strike that area would be South America. This is evident through our analysis of the data with the Analysis of Variance tests. We saw that there was a difference between the average meteorite mass for each continent. This difference was most notably seen in Antarctica and South America, but since we wanted to see where the largest meteorites struck, we chose to analyze South America even further because its mean mass was above the grand mean. Additional analysis of South America's meteorite mass saw that it was statistically significant and had a much larger mean mass than all the other continents. If you happen to want to travel or live in a different continent and are worried about meteorites you should not live in South America because they have the largest meteorites.

**References**

“7 Types of Data Bias in Machine Learning.” *Lionbridge AI*, 1 Dec. 2020, lionbridge.ai/articles/7-types-of-data-bias-in-machine-learning/.

“Meteorite Landings - Dataset by Nasa.” *Data.world*, 10 Mar. 2017, data.world/nasa/meteorite-landings.

**Appendix**

options(max.print=976257)  
meteor = read.csv('Meteorite-landings.csv')  
meteor = meteor[complete.cases(meteor[,5:9]),]  
coordinates = data.frame(lon = c(meteor$reclong), lat = c(meteor$reclat))

coords2country = function(points)  
{   
 countriesSP <- getMap(resolution='low')  
 #countriesSP <- getMap(resolution='high') #you could use high res map from rworldxtra if you were concerned about detail  
  
 # convert our list of points to a SpatialPoints object  
  
 # pointsSP = SpatialPoints(points, proj4string=CRS(" +proj=longlat +ellps=WGS84 +datum=WGS84 +no\_defs +towgs84=0,0,0"))  
  
 #setting CRS directly to that from rworldmap  
 pointsSP = SpatialPoints(points, proj4string=CRS(proj4string(countriesSP)))  
 #mass = points$mass..g.  
  
 # use 'over' to get indices of the Polygons object containing each point   
 indices = over(pointsSP, countriesSP)  
   
 # return the ADMIN names of each country  
 #indices$ADMIN   
 #indices$ISO3 # returns the ISO3 code   
 #indices$continent # returns the continent (6 continent model)  
 #indices$REGION # returns the continent (7 continent model)  
 #coords <- coords2country(coordinates)  
indices$mass <- meteor$mass..g.  
indices[order(indices$REGION),]  
}

coords <- coords2country(coordinates)  
table(coords$REGION)

africa\_index <- c(grep('Africa', coords$REGION))  
africa\_mass <- coords[africa\_index,-1]  
africa\_mean <- mean(africa\_mass$mass)/1000  
africa\_sd <- sd(africa\_mass$mass)/1000  
  
  
antarctica\_index <- c(grep('Antarctica', coords$REGION))  
antarctica\_mass <- coords[antarctica\_index,-1]  
antarctica\_mean <- mean(antarctica\_mass$mass)/1000  
antarctica\_sd <- sd(antarctica\_mass$mass)/1000  
  
asia\_index <- c(grep('Asia', coords$REGION))  
asia\_mass <- coords[asia\_index,-1]  
asia\_mean <- mean(asia\_mass$mass)/1000  
asia\_sd <- sd(asia\_mass$mass)/1000  
  
australia\_index <- c(grep('Australia', coords$REGION))  
australia\_mass <- coords[australia\_index,-1]  
australia\_mean <- mean(australia\_mass$mass)/1000  
australia\_sd <- sd(australia\_mass$mass)/1000  
  
europe\_index <- c(grep('Europe', coords$REGION))  
europe\_mass <- coords[europe\_index,-1]  
europe\_mean <- mean(europe\_mass$mass)/1000  
europe\_sd <- sd(europe\_mass$mass)/1000  
  
north\_america\_index <- c(grep('North America', coords$REGION))  
north\_america\_mass <- coords[north\_america\_index,-1]  
north\_america\_mean <- mean(north\_america\_mass$mass)/1000  
north\_america\_sd <- sd(north\_america\_mass$mass)/1000  
  
south\_america\_index <- c(grep('South America', coords$REGION))  
south\_america\_mass <- coords[south\_america\_index,-1]  
south\_america\_mean <- mean(south\_america\_mass$mass)/1000  
south\_america\_sd <- sd(south\_america\_mass$mass)/1000

pf(65.23, 6, 32011, lower.tail = FALSE) #p-value for the ANOVA Test  
  
pt(-7.395, 32011, lower.tail = TRUE)\*2 #p-value for the contrast test statistic