**STA502**

**The Analysis of UFO data**

**Lina Lee**

**Ⅰ. Abstract**

This dataset includes reports of UFO sighting from 1998 to 2014. The dataset includes 11 variables including the number of reports of UFO, country of sightings, duration of the sighting and properties of UFO sighted, and geographical location etc. I focus on analyzing the number of reports over the geographical location and analyzing the range of longitude and latitude for each country, and the duration of the sighting UFO in seconds for each country. I also test the difference in mean of the duration of the sighting in seconds between Canada and the US or between two countries in Europe and two countries in America.

**Ⅱ. Introduction**

This dataset has over 80,000 reports of UFO sighting from 1998 to 2014. The dataset has 11 variables: City in which UFO was sighted, State in which UFO was sighted, Country of Sighting, shape of the UFO, duration of the sighting in seconds, duration of the sighting in hours and min, sighting description, date posted, posted date of the sighting, latitude coordinate of the sighting, longitude coordinate of the sighting. The country contains various countries over different continents. Shape includes chevron, cigar, circle, cone, cross, fireball triangle, disk, egg, sphere, flash, cylinder, delta, diamond, etc. The duration widely spread from a few seconds to hours, and people reported in many ways. The range of Longitude is from -177 to 153, the range of latitude is from -42 to 70. My research question is listed below. Considering Latitude, longitude and duration are continuous variables, and country is a categorical variable. Therefore, I analyze value continuous variables by the category of country.

Research question

1. Is there any trend in the number of reports by groups of latitude and longitude?
2. Analysis for each country

-the range of latitude and longitude of each country?

-the number of reports for each country

-statistics for duration of the sighting UFO in seconds for each country

1. The mean of the durations is different by each country?

**Ⅲ. Methods**

I tried to analyze some trends in the number of reports of UFO depending on geographical criteria such as latitude and longitude. I divided latitude into 5 groups, and the range of each group is 23. The criteria are based on division lines of latitude, although only I used Tropic of Cancer(23° 26′) and Tropic of Capricorn(23° 26′) due to the censored range of latitude for the UFO data. Longitude was divided into 12 groups. The range of longitude is wider than the range of latitude. I generated a table of the number of reports by groups using Proc SQL labeling properly. When I divide groups using case when statement in Proc SQL. After generating groups for geographical coordinate, I counted the number of reports by counting the number of the posted date of the sighting. I used Proc Print to print table1 using label as an option. I also created plots putting the number of reports on the y-axis and having latitude and longitude on X-axis each using Proc Sgplot to see the result clearly.

I got the summary by each country using Macro function. I wanted to see the range of latitude and longitude of each country, statistics of duration (seconds), and the number of reports of sightings by country. I used Macro function because there are several countries that I need to analyze, so I can easily apply the same code to implement the same analysis by each country. I included the count of the number of the reports of UFO, average, minimum and maximum of the duration (seconds), and average, minimum and maximum of latitude and longitude using Proc SQL while restricting the data for a country using Where statement.

I also used Proc Glm and Contrast statement to test the difference of the mean duration of the sighting in seconds between Canada and the United States. I also tested whether there is the difference in mean duration of EU countries (the United Kingdom and Germany) and the countries in the America continent (Canada and the United States).

**Ⅳ. Results**

***Table1: the number of reports by group of latitude***

| *Obs* | *group by latitude* | *The number of reports of UFO* |
| --- | --- | --- |
| *1* | latitude<-23 | 897 |
| *2* | -23=<latitude<0 | 215 |
| *3* | 0=<latitude<23 | 833 |
| *4* | 23=<latitude<46 | 73203 |
| *5* | 46=<latitude | 10333 |

In the table1, most of the reports are reported at the area where latitude is between 23 and 46, the number is 73203. The second one is the area where latitude is greater than 46m the number is 10333. This result is related to the fact. I need to check the latitude range of the United State. An interesting point is that the area where latitude is between -23 and 0 has the least reports number.

Figure1



The number of reports in longitude between -90 and -60 is the largest. The number is 41124, which is dramatically different from the group of which range above -60. The second one is 28564 for the group of the longitude from -120 to -90. The groups in the range from 150 to -30 has quite smaller the number of reports than the one of -30 to -150. However, longitude between -30 and 0 stands out among them. By observing the scatter plot, it looks like regression line does not fit.

***Table2: Scatter Plot of the number of reports by group of longitude***

| *Obs* | *group by longitude* | *The number of UFO reports* |
| --- | --- | --- |
| *1* | 150=<longitude<180 | 383 |
| *2* | 120=<longitude<150 | 524 |
| *3* | 90=<longitude<120 | 328 |
| *4* | 60=<longitude<90 | 343 |
| *5* | 30=<longitude<60 | 332 |
| *6* | -30=<longitude<0 | 2629 |
| *7* | -60=<longitude<-30 | 181 |
| *8* | -90=<longitude<-60 | 41124 |
| *9* | -120=<longitude<-90 | 28564 |
| *10* | -150=<longitude<-120 | 11073 |

‘

**Figure2**



***Table 3. summary for the United State***

***Summary of latitude for country "us"***

| *The number of reports of UFO* | *The mean of latitude* | *The minimum of latitude* | *The median of latitude* | *maximum of latitude* |
| --- | --- | --- | --- | --- |
| 69954 | 38.40 | 17.97 | 38.98 | 70.13 |

***Summary of longitude for country "us"***

| *The mean of longitude* | *minimum of longitude* | *The median of longitude* | *maximum of longitude* |
| --- | --- | --- | --- |
| -95.40 | -149.90 | -89.99 | -65.83 |

***Summary of duration(sec) for country "us"***

| *The mean of duration(sec)* | *minimum of duration(sec)* | *The median of duration(sec)* | *maximum of duration(sec)* |
| --- | --- | --- | --- |
| 5301.85 | 0.00 | 120.00 | 66276000.00 |

In the table 3, the number of reports of UFO for the US is 69954. The range of latitude of the US is from 17.97 to 70.13. The range of longitude of the US is from -149.9 to -65.83. The minimum duration (seconds) reported is 0 and maximum is 66276000.00 seconds. The mean of duration (seconds) is 5301.85.

In the appendix, the table 4, the number of reports of UFO sightings in Australia is 693. The range of latitude of Australia is from -42.88 to -12.46. The range of longitude for Australia is from 114.25 to 153.61. The minimum duration is 0 and the maximum is 1209600 seconds, which is smaller than the maximum durations reported in the USA. The number of reports of UFO sighting in Australia is far smaller than the number of reports in the United States.

In the appendix, table 5, the number of reports of UFO sightings is 3266, which is greater than the one of Australia, but smaller than the one of United States. The range of latitude is from 41.97 to 72.70, and the range of longitude is from -138.67 to -52.67. The minimum of duration reported is 0 and the maximum of duration reported is 82800000 seconds, which is larger than the maximum of duration reported in Australia or the United State.

In the appendix, table 6, the number of reports of UFO in the United Kingdom is 1904, which is smaller than the number of the United States and Canada. The range of latitude is from 50.11 to 57.85. The range of longitude is from -7.20 to -0.01. The minimum duration reported is 0 and the maximum of duration is 97836000, which is greater than all other countries.

***Table 7: Test difference of the mean duration between some countries***

|  |
| --- |
| ***The GLM Procedure*** |

|  |
| --- |
| ***Dependent Variable: durationsec*** |

| *Contrast* | *DF* | *Contrast SS* | *Mean Square* | *F Value* | *Pr > F* |
| --- | --- | --- | --- | --- | --- |
| *H0:mu\_ca=mu\_us* | 1 | 1.394247E12 | 1.394247E12 | 3.84 | 0.0501 |
| *H0:mu\_EU=mu\_America* | 1 | 280351851058 | 280351851058 | 0.77 | 0.3796 |

The mean difference of duration in seconds (table 7) between Canada and the US is significant, which means the mean of duration in Canada is different from the one of the US significantly. The mean of the EU countries (United Kingdom, Germany) and the mean duration of the countries in America (Canada and the United States) are not different significantly.

**Ⅴ. Conclusion**

When analyzing by groups for latitude, the group range from 23 to 46 included the highest number of reports of UFO sightings. The range of the group corresponds to some parts of latitude in the US and the United Kingdom. Among groups of longitude, the group ranged from -90 to -60 has the highest number of the reports, the range of the group corresponds to the longitude range of the US. The US has the highest number of reports among countries as 69954. Considering the UFO reported most in the US, the result in the analysis of the number of UFO reports by groups of latitude and longitude respectively is consistent with the result in the analysis by country. The second highest country in the number of the reports is Canada. The highest maximum duration among countries is one of the United Kingdom. It was found that the difference in mean duration of the sighting in seconds between Canada and the United State, however, the result says there is no difference in mean duration of the sighting in seconds between the two countries in Europe and the two countries in America.

**Ⅵ. Reference**

National UFO Reporting Center. (2016). UFO sightings [complete.csv]. Retrieved from https://www.kaggle.com

/NUFORC/ufo-sightings

**Appendix**

***Table 4. The summary for Australia***

1. ***Summary of latitude for country "au"***

| *The number of reports of UFO* | *The mean of latitude* | *minimum of latitude* | *The median of latitude* | *maximum of latitude* |
| --- | --- | --- | --- | --- |
| 593 | -32.79 | -42.88 | -33.86 | -12.46 |

1. ***Summary of longitude for country "au"***

| *The mean of longitude* | *minimum of longitude* | *The median of longitude* | *maximum of longitude* |
| --- | --- | --- | --- |
| 142.94 | 114.25 | 146.81 | 153.61 |

***(c) Summary of duration(sec) for country "au"***

| *The mean of duration(sec)* | *minimum of duration(sec)* | *The median of duration(sec)* | *maximum of duration(sec)* |
| --- | --- | --- | --- |
| 3453.42 | 0.00 | 120.00 | 1209600.00 |

\*au=Austraila

***Table 5. The summary for Canada***

1. ***Summary of latitude for country "ca"***

| *The number of reports of UFO* | *The mean of latitude* | *The minimum of latitude* | *The median of latitude* | *maximum of latitude* |
| --- | --- | --- | --- | --- |
| 3266 | 47.24 | 41.97 | 46.04 | 72.70 |

1. ***Summary of longitude for country "ca"***

| *The mean of longitude* | *The minimum of longitude* | *The median of longitude* | *maximum of longitude* |
| --- | --- | --- | --- |
| -90.53 | -138.67 | -79.83 | -52.67 |

1. ***Summary of duration(sec) for country "ca"***

| *The mean of duration(sec)* | *minimum of duration(sec)* | *The median of duration(sec)* | *maximum of duration(sec)* |
| --- | --- | --- | --- |
| 26508.97 | 0.00 | 120.00 | 82800000.00 |

\*ca : Canada

***Table 6. The summary for United Kingdom***

1. ***Summary of latitude for country "gb"***

| *The number of reports of UFO* | *The mean of latitude* | *minimum of latitude* | *The median of latitude* | *maximum of latitude* |
| --- | --- | --- | --- | --- |
| 1904 | 52.82 | 50.11 | 52.55 | 57.87 |

1. ***Summary of longitude for country "gb"***

| *The mean of longitude* | *minimum of longitude* | *The median of longitude* | *maximum of longitude* |
| --- | --- | --- | --- |
| -1.86 | -7.20 | -1.62 | -0.01 |

1. ***Summary of duration(sec) for country "gb"***

| *The mean of duration(sec)* | *minimum of duration(sec)* | *The median of duration(sec)* | *maximum of duration(sec)* |
| --- | --- | --- | --- |
| 66006.68 | 0.00 | 120.00 | 97836000.00 |

\*gb : the United Kingdom