

Factors Affecting the Duration of UFO

Presented by: Xiruo Li

SINOZ



Data Description

- The data is about reports of UFO sightings around the world. It has 3000 cases and there are 8 variables for each case.

Variable	Type	Len
city	Char	27
country	Char	2
datetime	Num	8
duration	Num	8
latitude	Num	8
longitude	Num	8
shape	Char	8
state	Char	2

Data Description

- First 10 observations of data

Obs	datetime	city	state	country	shape	duration	latitude	longitude
1	05DEC12:00:00:00	north bay (canada)	on	ca	fireball	30	46.3	-79.45
2	08APR13:20:45:00	croydon (uk/england)		gb	circle	40	51.383333	-0.1
3	28OCT03:16:00:00	victoria (canada)	bc	ca	light	60	46.216667	-63.483333
4	01JUL01:20:00:00	london (uk/england)		gb	oval	900	51.514125	-0.093689
5	21JUL07:19:00:00	slave lake (canada)	ab	ca	disk	60	55.283333	-114.783333
6	24JUL06:23:10:00	pulborough (uk/england)		gb	light	40	50.95	-0.516667
7	27JUN09:23:00:00	stoke-on-trent (uk/england)		gb	circle	30	53	-2.183333
8	03APR10:12:00:00	punta gorda	fl	us	light	240	26.9294444	-82.0455556
9	27NOV05:03:00:00	lichfield (uk/england)		gb	light	120	52.683333	-1.816667
10	21OCT12:03:20:00	durham county (uk/england)		gb	light	600	54.766667	-1.566667

Goal of Analysis

- Using ANOVA (proc glm) to find what factors have significant contribution to the duration of UFO sightings
- Factors: country, shape, latitude, longitude



Data Manipulation

- Convert latitude, longitude from continuous variables into categorical variables (4 levels:1,2,3,4), according to the quantile.

latitude	latitude_level
[19.50, 42.00]	1
(42.00, 46.30]	2
(46.30, 51.83]	3
(51.83, 72.70]	4

longitude	longitude_level
[-157.72, -95.70]	1
(-95.70, -78.87]	2
(-78.87, -2.66]	3
(-2.66, 1.75]	4

- Drop the outlier of duration: 52623200 seconds

Slice of Cross Tabulation

proc tabulate:

				duration		
				Mean	Std	N
country	shape	latitude_level	longitude_level			
ca	changing	2	2	954.50	1402.04	10
			3	1816.50	2368.70	10
		3	1	653.61	441.17	7
			2	5400.00	.	1
			3	960.00	691.95	3
		4	1	150.00	127.28	2
	cigar	2	2	132.91	136.30	11
			3	555.33	1154.14	9
		3	1	5711.43	14154.90	7
			3	610.00	834.39	2
		4	1	9001.00	12726.51	2

It's clear that the data is unbalanced. So, **proc glm** is used in ANOVA.

ANOVA

Original model: duration=country+shape+latitude_level+longitude_level

Source	DF	Sum of Squares	Mean Square	F Value	Pr >
Model	20	1044193266.5	52209663.324	1.01	0.445
Error	2978	153825808430	51654065.96		
Corrected Total	2998	154870001697			

R-Square	Coeff Var	Root MSE	duration Mean
0.006742	694.0422	7187.076	1035.539

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	2	8085186.9	4042593.5	0.08	0.9247
shape	12	425408070.0	35450672.5	0.69	0.7663
latitude_level	3	174565009.1	58188336.4	1.13	0.3369
longitude_level	3	436135000.4	145378333.5	2.81	0.0379

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	2	115591756.9	57795878.4	1.12	0.3268
shape	12	432102388.4	36008532.4	0.70	0.7558
latitude_level	3	12738934.0	4246311.3	0.08	0.9697
longitude_level	3	436135000.4	145378333.5	2.81	0.0379

- The model is insignificant.
- Country, shape, latitude_level have high p values, drop them
- R square is low

ANOVA

Final model: duration=longitude_level

The GLM Procedure

Dependent Variable: duration

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	355926576.69	118642192.23	2.30	0.0754
Error	2995	154514075120	51590676.167		
Corrected Total	2998	154870001697			

R-Square	Coeff Var	Root MSE	duration Mean
0.002298	693.6162	7182.665	1035.539

Source	DF	Type I SS	Mean Square	F Value	Pr > F
longitude_level	3	355926576.7	118642192.2	2.30	0.0754

Source	DF	Type III SS	Mean Square	F Value	Pr > F
longitude_level	3	355926576.7	118642192.2	2.30	0.0754

- The model is significant at 0.1 significance level.
- The R square is low, though.

Multiple Comparison

adjust=Tukey

longitude_level	duration LSMEAN	LSMEAN Number
1	1553.55487	1
2	723.96804	2
3	1141.71247	3
4	721.36497	4

Least Squares Means for Effect longitude_level				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	829.586831	-123.824530	1782.998191
1	3	411.842406	-540.621395	1364.306207
1	4	832.189907	-123.136591	1787.516405
2	3	-417.744424	-1369.571773	534.082924
2	4	2.603076	-952.088878	957.295030
3	4	420.347500	-533.398166	1374.093167

For duration time, longitude level: 1>3>2≈4

Conclusion

- Longitude level has significant contribution to the duration of UFO sightings at 0.1 significance level.
- UFO has the longest duration time where longitude is in $[-157.72, -95.70]$.

