|  |  |
| --- | --- |
| * Arnold Reyna * DSC 530 Final Project – 250 word summary: | |
|  | |
|  | |
|  | |
| **Hypothetical question:** |  | |

**On days there is a full moon, does crime increase**?

Full moons have been associated with affecting behavior such as suicide, sleepwalking and violence. This is known as the Lunar effect. This belief goes back centuries and has roots even in the name the name: Latin word for moon is Luna, from which we get the words 'lunacy' and 'lunatic'. There is some proof that the stronger lunar gravitational pull has influence on the physiology of humans. Some counter-theories point to time before modern lighting- Light of a full moon kept people up at night, leading to sleep deprivation that could have caused some of the issues witnessed.

**Outcome of my EDA:**

Reviewing data over 2 years of crime in Austin Texas, showed there was not statistical evidence that crime increased on dates there were Full moon vs non Full moon dates. Only some boarderline evidence was found on observing some graphs (see powerpoint):

* Statistics on Severity of crimes (1-5(highest)):
* crime type mean non full moon : = 1.5308250546043825
* crime type mean full moon : = 1.511050794881737
* Difference between Means is : **-1.97743 %**
* crime type Variance non full moon : = 0.9361059155781544
* crime type Variance full moon : = 0.9078491281403448
* Difference between Variance is : **-2.82568 %**
* crime type Standard Deviation non full moon : = 0.967525666625002
* crime type Standard Deviation full moon : = 0.9528111712927934
* difference between Standard Deviation full moon vs non **= -0.01471**

Standard deviation difference is only 0.0147- too small to be significant. (To put that in perspective, the difference in height between men and women is about 1.7 standard deviations)

Coef of 'Severity' is negative, which indicates Full moon has opposite effect, but is a very small #.

The **P-value 0.3** which shows to be greater than .05, so are statistically insignificant.

**Coef of 'Severity' is negative**, which indicates Full moon has opposite effect, but is a very small number so not significant.

The coefficient of determination**, R2 value** for this model is small (.000047), which means that moon doesn’t account for a substantial part of the variation. R2 does not apply to logistic regression. Since I used Full moon as a logistical attribute, this is the case here.

**Overall:** There isn’t strong enough evidence in the data to indicate the moon influences crime.

**o What do you feel was missed during the analysis?**

Since this is Logistical analysis, I didn’t have another strong numeral variable to compare to so I could determine a link. For example, coefficient of determination, R2 does not apply to logistic regression, only for comparing to other variables. Some other variables which should have been considered for this project crimes are:

* Range of days around full moons- it would be helpful to compare crime increases on dates around full moon dates since full moons span over a few days
* Factoring out crime spikes on weekend which occur regardless of full moon.

**Were there any variables you felt could have helped in the analysis?**

* **Gender**- comparing only men or women on full moon vs non full moon dates
* **age of criminal**- To compare spikes in crime based on same age.
* **GPS coordinates**- would be good to geographically isolate areas affected by the moon, instead of the whole city.

**Were there any assumptions made you felt were incorrect?**

Crime report dates: I assumed the date taken on report is the day of the crime. Could be date the report was made.

Crimes that occurred after midnight were counted as the next day, which could be missing full moon date on data.

**What challenges did you face, what did you not fully understand?**

I feel there were so many choices on how the data can be analyzed from Chapters 10 and 11. The challenge was finding the right analytical techniques which are meaningful for the data selected.