Ines Belkhodja

Investigating Trojan Asteroids

1. Background

Our Solar System contains millions of small bodies orbiting. For example, there are different types of asteroids that co-orbit various planets, in particular Trojan Asteroids that co-orbit Jupiter. From Jet Propulsion Lab's Solar System Dynamics for Small Body database, I examined multiple physical characteristics of these Trojan Asteroids, as the sole mission and purpose of this online database are to "focus on determining the motion and physical parameters of natural planetary objects and our primary products, services, and charter". This large group is divided into two populations of asteroids, which was the first interesting fact that steered my investigation into how these two populations are different, shown below in Image 1. The two populations are called L4 and L5 Trojans. L5 and L4 stand for Lagrange points 4 and 5, which are positions in space where objects sent there tend to stay put and are stable, shown in Image 2. The L1, L2, and L3 points are unstable regions; hence only extremely rare occurrences where these natural objects and other small bodies occupy these locations.

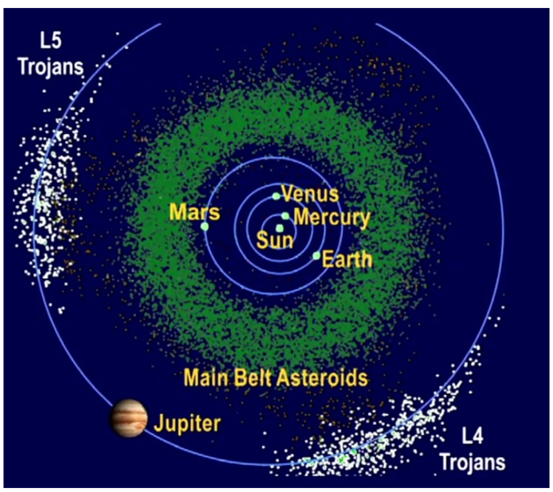

Image 1. L4 and L5 Trojan Asteroids

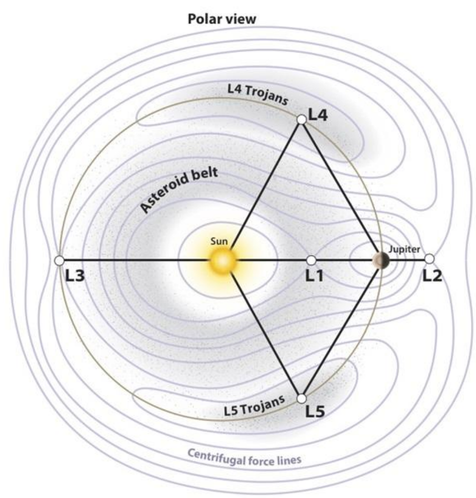

Image 2. Jupiter Lagrange Points

By knowing these two populations exist, my report dives deep into how various physical characteristics of these asteroids correlate with the two populations. Therefore, more specifically, I explored variables where these are common characteristics that people understand and ponder. Hence, I chose diameter, inclination, eccentricity, Jupiter MOID, and mean anomaly for the Trojan Asteroids. My investigation question relied on diameter because I thought it was an interesting and common variable to look at, so my question was "How does diameter affect eccentricity, inclination, and Jupiter MOID with the two populations from mean anomaly?" Diameter (km) looks at a straight line passing from side to side through the center of a body or figure. Inclination (deg) is the angle between the axis of the Trojan asteroids and a fixed reference angle. Eccentricity (no units) is the deviation of a curve or orbit from circularity; eccentricity equal to 0 is a perfect circle. Jupiter's minimum orbit intersection distance (AU) is the minimum distance between the orbits of two objects. In other words, the minimum distance of the asteroid's orbits away from Jupiter. Last, the mean anomaly (deg) is the angle between lines drawn from the Sun to the perihelion B (Jupiter in my case) and to a point (Trojan asteroids) moving in the orbit at a uniform rate corresponding to the period of revolution of the planet, shown as the green line in the Image 3 below. I showed the GIF in my presentation but it will be an image in my final investigation.

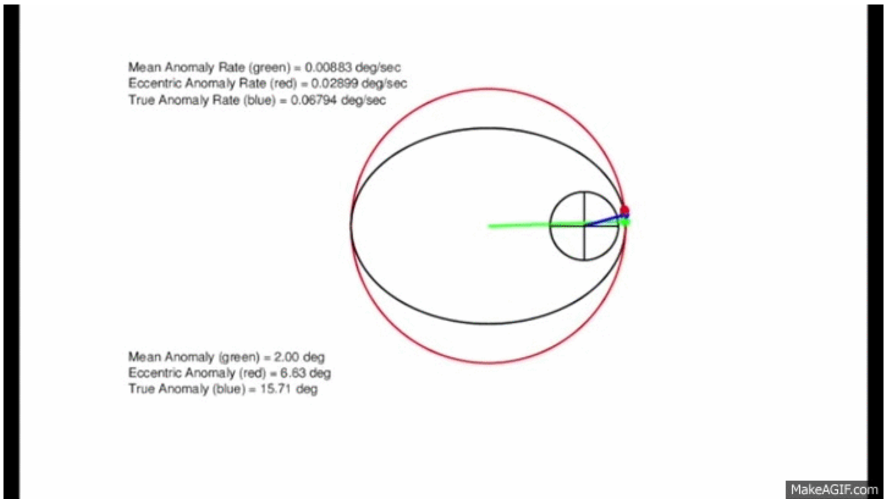

Image 3. Mean Anomaly (green) understanding for division of two Populations with Eccentric Anomaly And True Anomaly

JPL's database has sources of error or uncertainty for each of my chosen variables, for example, the column's name for error in inclination is called "i-sigma". To keep in mind throughout this report, the bias in this database is that there are many Trojan asteroids with unknown diameters, which will cause an imbalance in plotting my data with other chosen variables. This source of bias comes from the difficulty of measuring the diameter of these small bodies that are very far away from us.

2. **Procedure**

Gathering the data from JPL's database was not complex. I first selected "Jupiter Trojan" under "Limit by Orbit Class". Then, for the "Output Selection Controls", I selected only the column with the variables I wanted to look at more in-depth which I covered in the background section. After, I download the CSV file which was not too big as there are about 12,500 rows, which is the total amount of Trojan asteroids. I utilized all rows because I believe that my results will have a stronger conclusion rather than using a random sample of these asteroids. The first immediate step was to plot these variables to histograms and notice any relationships in my Jupyter Notebook using Python. I used various imported packages to start my code analysis such as pandas, and matplotlib. After plotting each variable on histograms, the most significant histogram was about the mean anomaly because it helped me advance my investigation forward. I first read my CSV file using pandas through the method pd.read_csv(). To plot my histogram, I used plt.hist().

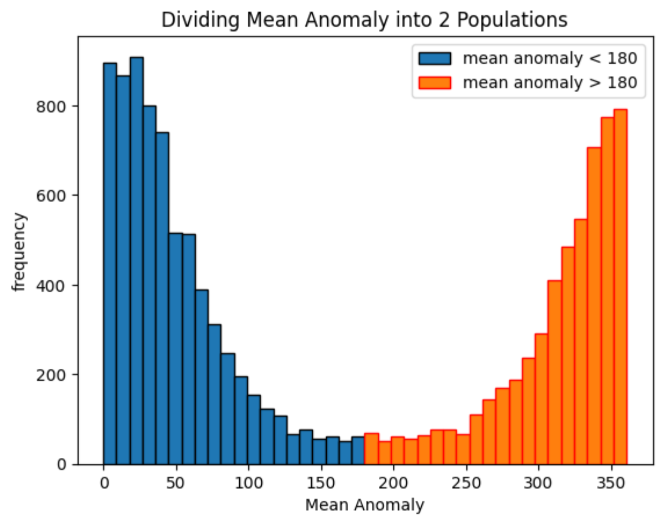

Figure 1. Dividing Mean Anomaly into 2 Populations

From the figure above and my understanding of mean anomaly from Image 3, mean anomaly does a full cycle of 360 degrees, and with the two populations on either side of Jupiter, I divided my histogram into two equal populations of less than 180 degrees and more than 180 degrees. When continuing with further investigations, I look at each variable with the 2 populations of under and over 180 degrees, shown in Figure 1. This revelation is the driving force behind my further investigation of how the L4 and L5 Trojan asteroids similar or different. I noted that there are asteroids in the middle, so at 180 degrees. However, there are no Trojan asteroids that are in that area, so these frequency numbers could be very small debris floating in those regions that co-orbit Jupiter.

Therefore, my next step was to include all my variables investigated in one scatter plot to visualize the initial trends in my data. I used a log scale on my data to get a clearer trend.

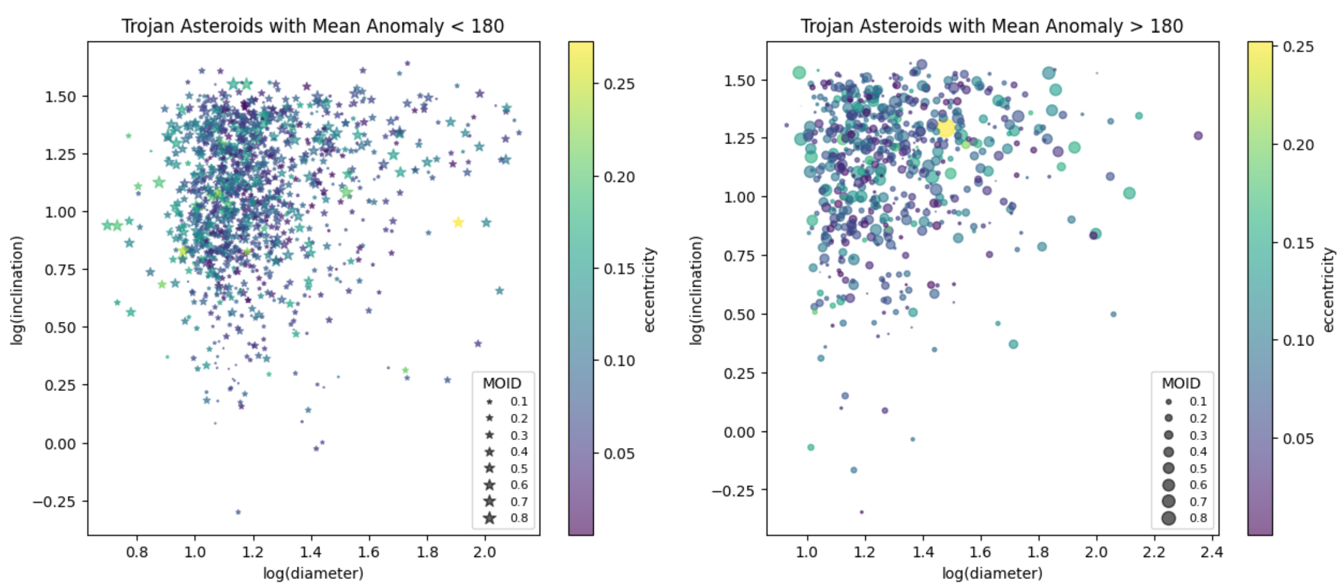

Figure 2. All four variables of diameter, inclination, eccentricity, and Jupiter MOID for the mean anomaly less than and greater than 180

I can see some weak correlations with my data. First, there is a proportional relationship between diameter and inclination; hence as the log diameter increases, the log of inclination increases as well for both populations. Eccentricity does increase with higher diameters, but an estimate of a weak correlation is due to the population with a mean anomaly > 180. Jupiter MOID is hard to evaluate as the different sizes of my markers are closely increasing together. Furthermore, in my code, I decided to multiply each of my Jupiter MOID values with 100 because they were too small to fully grasp the differences in size. Therefore, with extremely inadequate conclusions, I need to proceed in my investigation by looking more in-depth at the statistical calculations that I learned throughout this semester with Python.

For the rest of the procedures in my investigation, I plot scatterplots with each of my variables to see the various relationships I have among all of them. I used plt.scatter() for my scatter plots. Next, I used optimization to fit a linear regression model to my scatter plots of diameter with inclination and eccentricity through code learned in Lab 4. I wanted to see how strong a straight line fits my data and compare the slopes and correlations in both populations of my asteroids, so the mean anomaly is less than and greater than 180 degrees. The SpearmanR correlation coefficient is calculated for both populations using the built-in method in Python. I used it to statistically represent the relationship between diameter and inclination for these two variables, and I chose Spearman, as a change from my presentation where I used PearsonR correlation coefficient, in my investigation because one of its assumptions does not rely on the data to be normally distributed. It is evident in my histograms that my data is not normally distributed. The stats.spearmanr() method will allow me to do so. Following these calculations, I will do hypothesis testing as a next step. The null hypothesis always states there is no relationship between the measured phenomenon (the dependent variable) and the independent variable. To reject my null hypothesis, the confidence level in my investigation has to be 95 percent, so the probability value (p-value) has to be greater than 0.05 after going through my significance test. If the p-value is less than 0.05 from this test, then it will allow me to conclude if I can reject the null hypothesis, if not then I fail to reject the null hypothesis. Furthermore, I plotted a residual model for my model fit line regression to visualize the error with my line of best fit, as they are an important tool to understand how strong the trend is. To calculate the residuals, I subtracted the data value at a certain x point with that x value on the model fit line. Following the residual plots, I also calculated the residual sum of squares, which is a measure of the level of variance in the error term, or residuals, of a regression model. This statistical representation removes the negative results as some points may fall below the model fit line. Therefore, as the value of the residual sum of squares gets closer to zero, then the model progressively is a perfect fit line. By looking at the residuals and the statistical representation, I will be able to conclude if a linear regression model fits my data well.

Finally, the last step in my investigation is to look at a Monte Carlo simulation to verify if the two populations are similar or not in diameter. I chose diameter as there are only 1879 values in the database which are not NaN. I used the np.isnan() to filter out all the NaN values and take them away. As diameter has only a small data size compared to my entire population of over 12,500 asteroids that include all values for inclination, eccentricity, and Jupiter MOID values, I decided to simulate more data to question whether or not the two populations of diameters are identical. I first used the np.random.multivariate_normal() learned in Lab 11 to simulate over 50 data points, by inputting it in a function that calculates an array of all the diameter means by doing it 10,000 times. With this large amount of time calculating the mean, I will have a better accurate representation of the diameter means in both populations. I did this step twice due to my population of diameters that have a mean anomaly smaller than 180 and the other population with a mean anomaly greater than 180. I plotted the array of diameter means into a histogram, using the plt.hist() method, and compared them to the true means of my populations. I calculated those two means in the beginning by using np.mean() for both populations. To complete my analysis with the Monte Carlo method, I carried out a T-test using the stats.ttest_ind() method, where it compares both the population means of my diameters, as they are two independent samples of my data. This test assumes that the two sample distributions are identical, which is how I formed my null hypothesis to be that the means for both populations are identically distributed. Hence, the alternative hypothesis is that there is a difference between both of them.

3. Analysis

My first step mentioned in the procedure was to look at some trends between the x-axis and the y-axis. I focused mainly on the diameter on the x-axis, with the changing variables for the y-axis. The scatter plots are shown below with the residuals beside them.

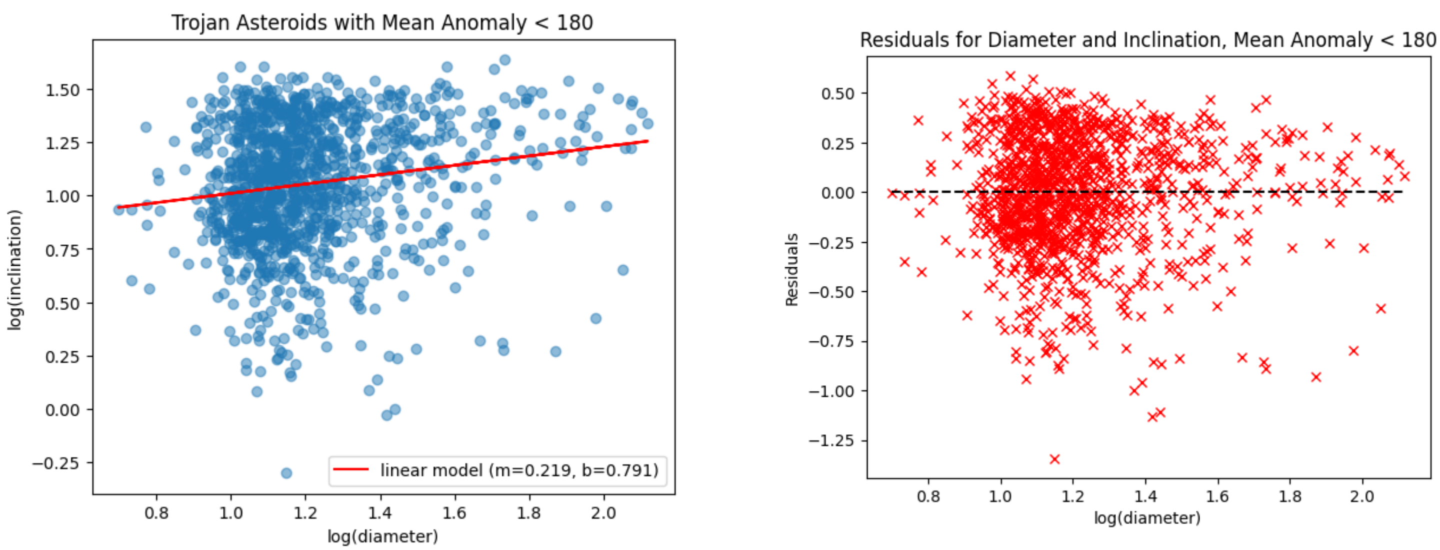

Figure 3. Diameter vs Inclination Scatter plot for Mean Anomaly less than 180 with residuals in the right plot

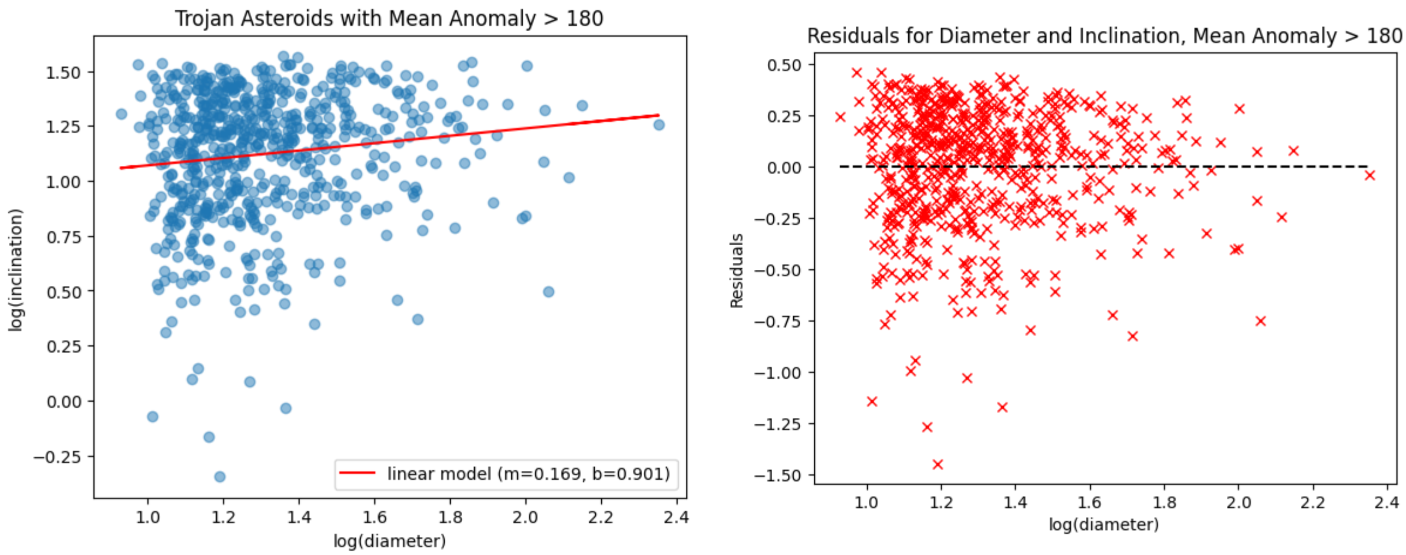

Figure 4. Diameter vs Inclination Scatter plot for Mean Anomaly greater than 180 with residuals in the right plot

From my statistical calculations using Python, I wanted to focus on diameter and find out whether there are correlations with the other three variables, eccentricity, inclination, and Jupiter MOID. First, with my SpearmanR correlation for diameter vs inclination for mean anomaly less than 180, I get a statistics value of 0.174, and a p-value of $8.112 * 10^{-10}$ so p-value << 0.05. For the other population greater than 180, I get a statistical value of 0.134 and a p-value of 0.0006. With these values received, I can do my hypothesis testing of whether there is a relationship between diameter and inclination for both populations. My null hypothesis is there is no relationship between my variables; hence, my alternative hypothesis is that there is a relationship between my variables. As the p-value is less than 0.05 from my reasoning in the procedure section, I can reject the null hypothesis in both my populations. Therefore, there is a relationship between the diameter and inclination of these asteroids. However, this correlation is not strong. Looking at the slope shown in Figure 3, the slope is equal to 0.219, while on the other hand, the slope in Figure 4, the slope is equal to 0.169. As these values are close to zero but positive, there is a weak and increasing relationship between my variables. In both of these figures, there is a residual plot shown, and it can be seen that there is a lot of error between my data point and the model fit line. Furthermore, the calculated residual sum of squares is equal to 105.22 and 54.92, respectively. These numbered values tell me that as they are far away from 0 (perfect model fit line), the linear regression line clearly does not fit my data points. To connect back to science, there is little to no research about the diameter and inclination of asteroids that I could currently find. Therefore, I realize how these trends are not comprehended yet by astronomers, but it would be interesting to finally understand my findings in a few years with better technology, and better branched out and financed research.

Additionally, there is a varying but interesting correlation between diameter and eccentricity for both populations, done in my code but not shown in the report. There is a relationship between diameter and eccentricity for the asteroids with a mean anomaly less than 180, but no relationship with a mean anomaly greater than 180 after concluding by doing a hypothesis test. I find the p-value for less than 180 to be equal to $2.459 * 10^{-13}$, so this number infers that I can reject the null hypothesis and conclude that there is a negative relationship for this chosen population of diameter and eccentricity. The slope for this population is equal to -0.169. As it is negative, there is an inverse relationship between diameter and eccentricity. As the diameter increases, the eccentricity slightly decreases, which is a trend I would have not guessed right away. On the other hand, the p-value for greater than 180 is equal to 0.49. Therefore, as this latter p-value is greater than 0.05, it allows me to conclude that I fail to reject the null hypothesis of no relationship. Also, this no-relationship is evident that the correlation between diameter and eccentricity for the population greater than 180 is extremely weak, with a slope of 0.0686. This revelation puzzles me because I am not sure why for one population there is a trend, while for the other population, there is no clear correlation. The residual plots do show my data points are far away from the model fit line, with a calculated residual sum of squares equal to 77.05 and 59.81, respectively. Therefore, these concluding correlations for both populations exhibited interesting results that I would have not known right away from Figure 2.

To my surprise, there is no relationship between diameter and Jupiter MOID, done in my code, but not shown as a figure in my report. I found the p-value for both populations to be over 0.05, so I obtained 0.69 and 0.36 for mean anomalies less than and greater than 180, respectively. Therefore, I can conclude that when the diameter increases, there is no further or nearer distance away from Jupiter for both populations. This finding challenged the way I was thinking because it was also a question asked during my presentation. I thought that there would be a trend where if the asteroid is larger, then it is maybe further away from Jupiter. However, that is not the case, where my interpretation and intuition are right. From this statistical revelation, I would have not guessed that there is no effect on these two variables. The minimum distance away from Jupiter is randomly distributed against its diameter.

As mentioned in the procedure section, I did a Monte Carlo simulation with my diameter values. There is a total of 1,879 diameter values with over 12,500 asteroids in the database. Therefore, I wanted to simulate 10,000 values for diameter and compare them with my different populations. I also did a null hypothesis testing, where my null hypothesis is that the two populations are identical/similar. Hence, the alternate hypothesis is that the two populations are different from each other. Getting insight from Lab 11, I followed some similar steps with functions and code, described in the procedure section. My results turned out to be different from my presentation, due to some code error, which I told in my presentation that the two populations are similar. I have found that the p-value is greater than 0.05, which fails to reject the null hypothesis. However, it is indeed incorrect. After careful reconsideration of my code and the t-test with my two independent samples, I conclude that the p-value is 0 as the two populations are quite different with the diameter means, shown below in Figure 5.

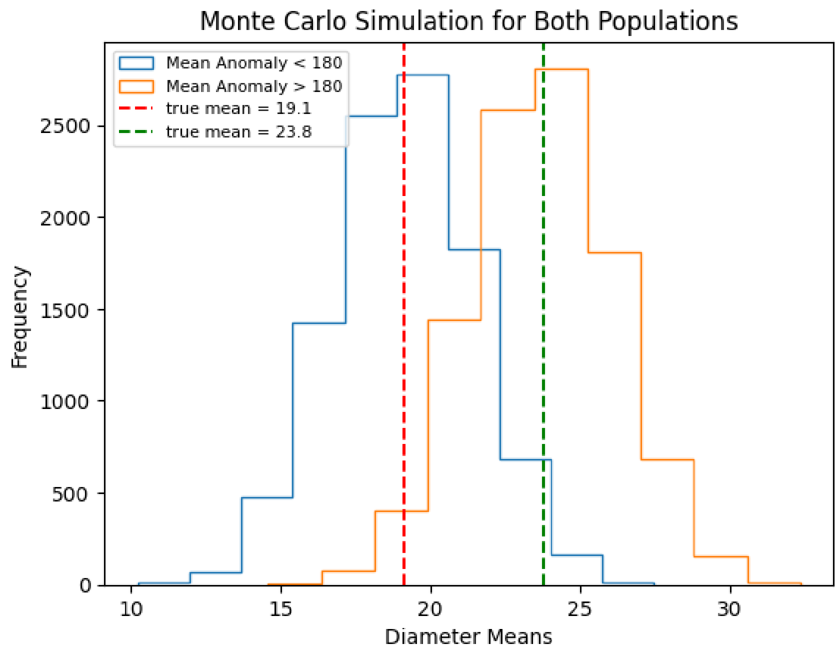

Figure 5. Monte Carlo Simulation for Diameter less than and greater than 180 degrees

Therefore, with my lack of diameter values, I can finally see how with more added values, I can visualize the difference in populations with simulated data, which is why a Monte Carlo simulation is very effective. This revelation was also fascinating to find out due to how the diameter is quite different with the various locations where these asteroids co-orbit with Jupiter. There is not much scientific writing and papers about Trojan asteroids to explain this phenomenon of not having identical means of diameters for both populations; nevertheless, it is significant to realize that these particular rocks in our Solar System are not sculpted to be all the same. Hence, more research is needed to accomplish my investigation and understand this scientific phenomenon to its fullest.

4. Conclusion

Throughout my investigation, I wanted to see how diameter is affected by other variables chosen, which were inclination, eccentricity, Jupiter MOID, and mean anomaly. Given how symmetrical the mean anomaly is, I divided the histogram into two populations of less and greater than 180, and this allows me to compare and contrast the two Trojan asteroid populations. Hence, with scatterplots and residual plots, I can conclude that diameter and inclination have a positive but weak correlation for both populations. Next, I found out that diameter and eccentricity have a negative and very poor correlation for the population of mean anomaly less than 180, and no clear relationship with mean anomaly greater than 180. Additionally, I examined how there is no trend at all for diameter and Jupiter MOID. Lastly, as diameter have a small amount of data points compared to the entire database, I looked into how a Monte Carlo simulation will enable me to further and efficiently conclude how diameter is compared in the two populations. Therefore, I can understand that the two populations of diameters are not identical after my statistical analysis. I am very interested in why these results happen in our Solar System, but unfortunately, there are limited papers on why these phenomena occur. NASA launched a space mission called Lucy in 2021 that will arrive near both Trojan asteroid populations in 2027. As a result of this new mission, I believe after several years, more and more papers will help me explain and verify why the findings of my variables happen to have interesting relationships.

There are a few limitations to my dataset as mentioned above, the diameter is not recorded for every asteroid in the full database. Therefore, there is an extent to my trends and relationships when I compared the diameter to the other variables. Luckily, the Lucy space mission will allow me to expand my work and eliminate a few of the limitations. Furthermore, this database was established from an Earth-based telescope. Hence, there is a bias towards larger asteroids that can be seen far away. This observational bias and limitations are not overlooked in my investigation and hope to have more information from the Lucy mission when it will start gathering data in the late 2020s.

For future work and expansion on this current exploration, I hope to include work error bars in my scatter plots to help with the linear model fit. I noticed with my residual plots and statistical analysis that there is a lot of error concerning my model fit line. Therefore, with errors in my plots, I believe I will have a stronger fit line using weighted linear regression. Moreover, this investigation only looked at 5 variables. However, the JPL database had over 55 columns of data, for example, the B-V index, absolute magnitude parameter, geometric albedo, etc. Thus, with more time, I want to examine and investigate the rest of these columns and find new relationships and trends unknown to men right now. In a few years, it would be very interesting to see some phenomenon backed up by science to help me explain my new findings and answer the big question of “why” it is happening.

Finally, something that I would have done differently is start my investigation in another way. I dived into my investigation with a question already in mind where the diameter was the center of it all. However, it limited my exploration because when I was finding the other variables, I was thinking, “Is there a possible trend with diameter?”. For that reason, I want to start a new investigation where I look at the relationships between all the known variables and then move forward with that revelation. However, I am still proud and curious to learn more about what I found in this investigation and apply my newfound knowledge to other asteroids in our Solar System, such as the ones with Mars. I want to see if the trends are identical or different if we look at other asteroids that co-orbit their planet, such as Earth or Venus. My findings at the moment have opened my eyes to more new paths that I could explore.

Works Cited

Anne Marie Helmenstine, Anne Marie. “What Is the Null Hypothesis?” *ThoughtCo*, 4 Sept. 2019, www.thoughtco.com/null-hypothesis-examples-609097.

Barone, Adam. “Residual Sum of Squares (RSS): What It Is, How to Calculate It.” *Investopedia*, 9 Oct. 2022, www.investopedia.com/terms/r/residual-sum-of-squares.asp#:~:text=The%20residual%20sum%20of%20squares%20(RSS)%20measures%20the%20level%20of,your%20model%20fits%20your%20data.

Davis, Joel. “Exploring Jupiter’s Trojan Asteroids.” *Astronomy.Com*, 12 Sept. 2018, astronomy.com/magazine/news/2018/09/exploring-jupiters-trojan-asteroids.

“Lucy.” *NASA*, 2 Mar. 2023, solarsystem.nasa.gov/missions/lucy/in-depth/#:~:text=Lucy%20is%20the%20first%20space,orbit%2C%20the%20other%20trailing%20behind.

“Small-Body Database Query.” *NASA*, ssd.jpl.nasa.gov/tools/sbdb_query.html#!#results. Accessed 1 May 2023.

“What Does SSD Do?” *NASA*, ssd.jpl.nasa.gov/about/. Accessed 1 May 2023.

“What Is a Lagrange Point?” *NASA*, solarsystem.nasa.gov/resources/754/what-is-a-lagrange-point/. Accessed 5 May 2023.