DISTANCE TO EXOPLANET STAR HYPOTHESIS TESTING (*Kepler Space Telescope*)

1. Background

The Kepler space telescope is deployed above the Earth (just like Hubble Space Telescope) which is tasked to find exoplanets – planets that orbit other stars. There are established methods for a telescope to detect such exoplanets. But before it can do so, it must point at a particular star. It then makes observations to detect if the star has a planetary system around it or not.

1. Hypothesis

The distance to the star (host to an exoplanet) is NOT similar across detection methods (which would indicate that either all discovery methods can be applicable to stars at any distance or may indicate that a particular discovery method could be beneficial for certain distances)

1. Data
2. Data source

https://www.kaggle.com/muhakabartay/markmarkohkeplerconfirmedplanets

1. Data description

The dataset is a conglomerate of many exoplanets discovered by eclectic methods. With each exoplanet, there are many fields recorded like orbital period, eccentricity etc. It also contains some information on the star each exoplanet may be revolving around.

1. Study Data

The complete dataset contains 3373 entries with certain proportion (out of whole) of exoplanets under each discovery method (appendix A – figure 1). But the number of entries in the classes were highly skewed/unbalanced. Hence, 5 methods of discovery were dropped. This rendered 3351 entries. Further, some distance-to-star readings were missing and were removed (1063)(appendix A – table 1). This yields a total of 2288 entries. But the final number of entries in the *transit timing variations* methods dropped too much. Hence the whole method was eliminated as well. The final dataset had only 3 methods (appendix A – figure 2) with a total of 2283 entries.

1. Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Classes in dataset | Imaging Method | Radial Velocity Method | Transit Method |
| Normality Testing | K-S Test Significant | K-S Test Significant | K-S Test Significant |
| Test Applicable | Kruskal-Wallis Test (non-parametric test for non-normal data) | | |
| Test Statistic value | 1021.896 ( *df = 2* ) | | |
| p-value | 0.000 | | |
| Significance level | 0.05 | | |
| decision | Reject null hypothesis | | |

1. Conclusion

The non-parametric test shows that the median distance to parent star of exoplanet (or the distributions under each category) is STATISTICALLY SIGNIFICANT (or the medians are not same) across the methods of discover by the Kepler Space telescope ( *p=<0.01 ; alpha = 0.05)*

This requires us to do further investigation (post-hoc analysis) to see if there is significance in pair-wise comparisons (pair-wise Wilcoxon Rank Sum Test)(number of pairs = 3).

Appendix A

Figure 1:

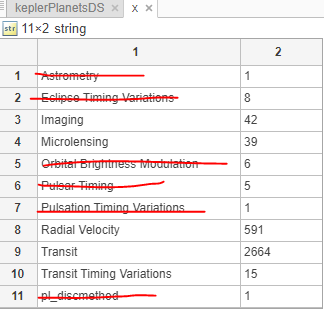
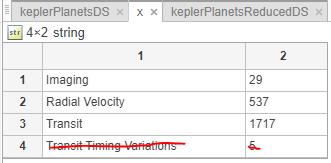


Table 1:

|  |  |
| --- | --- |
| Total Exoplanets | 3372 |
| Removal due to less entries in method of discovery | 21 |
|  | = 3351 |
| Removal of entries due to missing distance readings | 1063 |
|  | = 2288 |
| Removal of all transit time variation entries | 5 |
|  | = 2283 |

Figure 2:



Then individual datasets were created except for transit time variations. So, we have 3 datasets.