1. Abstract

The increasing number of exoplanet detections shows that planetary systems around other stars are common in the Universe but also that they may possess a wide range of physical and orbital properties. Physical and statistical models are needed to explain the relation between exoplanets and the characteristics of their host stars. In this work we analyze this issue through the application of multivariate statistical techniques.

The results did not indicate any of the exoplanet property can separate them into groups (or clusters). Further analyses must be conducted since we could not find a linear relationship between the exoplanet and its host star using the transformed data.

2. Some remarks about exoplanets

An extrasolar planet, or exoplanet, is a planet which orbits a star other than our Sun. As of May 2011, there are more than 500 confirmed exoplanets mainly detected by Radial Velocity (RV) and Transits (T).

**IMPORTANCE:** Their physical ($m_r$ and $r_i$) and orbital ($T$ and $e$) properties can help us to understand better the origin and evolution of the giant planets in the Solar System.

Levinson et al. (2010) have found evidence that comets are possibly of extrasolar origin.

**STATISTICAL CHALLENGES:** Represent a challenge in multivariate analysis (missing data, repeated measurements and outliers).
Finding “correct” relations among several variables.
Variables: different units, range of values.

**THE GOAL**

ANALYZE THE EXOPLANET AND HOST STAR DATA BY MEANS OF MULTIVARIATE STATISTICAL ANALYSIS, TO FIND A POSSIBLE RELATION BETWEEN THE SET OF VARIABLES.

3. Exoplanet Data Explorer, EDE (updated Wright et al. 2011)

<table>
<thead>
<tr>
<th>Name</th>
<th>Method</th>
<th>T (days)</th>
<th>a (AU)</th>
<th>e</th>
<th>min_r (K)</th>
<th>m_r (M)</th>
<th>RV (Km/s)</th>
<th>logg</th>
<th>logAM (d)</th>
<th>d (pc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7250</td>
<td>RV</td>
<td>2.21</td>
<td>0.47</td>
<td>0.80</td>
<td>0.76</td>
<td>0.24</td>
<td>-6.18</td>
<td>4.08</td>
<td>4.03</td>
<td>1.05</td>
</tr>
<tr>
<td>6671</td>
<td>RV</td>
<td>2.21</td>
<td>0.47</td>
<td>0.80</td>
<td>0.76</td>
<td>0.24</td>
<td>-6.18</td>
<td>4.08</td>
<td>4.03</td>
<td>1.05</td>
</tr>
<tr>
<td>HD 7444</td>
<td>RV</td>
<td>9.0000</td>
<td>0.20</td>
<td>0.47</td>
<td>0.80</td>
<td>0.76</td>
<td>0.24</td>
<td>-6.18</td>
<td>4.08</td>
<td>4.03</td>
</tr>
<tr>
<td>HD 7897</td>
<td>RV</td>
<td>9.0000</td>
<td>0.20</td>
<td>0.47</td>
<td>0.80</td>
<td>0.76</td>
<td>0.24</td>
<td>-6.18</td>
<td>4.08</td>
<td>4.03</td>
</tr>
</tbody>
</table>

**LSTAR = dSTAR (dSUN)**

Let A a set of p variables X and B a set of q variables Y. We are interested in linear relationships between A and B as well as interrelations among the X’s variables in A.

**Principal Component Analysis (PCA):** For $Z=x_\alpha x_\beta ... x_q$, where $a_\alpha ... a_q$ are chosen in such a way that Var($Z_\alpha$) is maximal, $\alpha_\beta ... a_q$, $Z_\alpha$ is the 1st PC. For the 2nd PC, an extra condition must be fulfilled: Cov($Z_\alpha$, $Z_\beta$) = 0. The number of PC to be chosen should explain about 85% of the total variance.

**Canonical Correlation Analysis (CCA):** We are interested in finding r (r = min(p, q)) linear relationships of the type: $U = a_1 x_1 + a_2 x_2 + ... + a_q x_q$. $V = b_1 y_1 + b_2 y_2 + ... + b_q y_q$ such that (Hotelling, 1936): 1) $\rho(U, V)$ is a maximum, and 2) $\rho(U, V)$ is a maximum and $\rho(U, U) = \rho(V, V) = 0$ ($p$ is the correlation).

From our data we obtain:

**Multivariate Analysis of the Variance (MANOVA):** Requires MVN and $\Sigma = \Sigma_{\mu} = \Sigma_{\nu} = \Sigma_{\omega}$

6. References

Hotelling, H. (1936), Biometrika, 28, 139.

**ACKNOWLEDGEMENTS**

I thank to the Schlumberger foundation and Consejo Nacional de Ciencia y Tecnología (CONACyT) for their financial support for this research. This research has made use of the Exoplanet Orbit Database and the Exoplanet Data Explorer both at exoplanets.org.
I also thank to professors Jagesh Babu and Donald Richards from the Department of Statistics (PSU) for their fruitful discussions.