Eurasian Academy of Sciences
Eurasian Econometrics, Statistics & Emprical Economics Journal
2021 Volume:18 S: 100- 113
Published Online January 2021 (http://econstat.eurasianacademy.org)
http://dx.doi.org/10.17740/eas.stat.2021-V18-07

DETERMINING THE DIFFERENCES OF EXOPLANETS ACCORDING TO THEIR CLASSES BY KRUSKAL WALLIS METHOD

Ünal Halit ÖZDEN * Kubilay ERİŞLİK**

**Prof. Dr., Istanbul Commerce University Faculty of Humanities and Social Sciences, Statistics Department, uozden@ticaret.edu.tr, ORCID ID: 0000-0003-0924-4848
**Research Assistant, Istanbul Commerce University Faculty of Humanities and Social Sciences, Statistics Department, kubilayerislik@ticaret.edu.tr, ORCID ID: 0000-0002-0744-4435

Received Date:08.11.2020, Revised Date:18.12.2020, Accepted Date:06.01.2021

Copyright © 2021 Ünal Halit ÖZDEN, Kubilay ERİŞLİK. This is an open access article distributed under the Eurasian Academy of Sciences License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

More than 2600 exoplanets have been confirmed thanks to the Kepler spacecraft that NASA sent in 2009. The confirmed exoplanets are classified by their parent stars. In this study, it is aimed to determine which properties of the exoplanets, which are classified by their main stellar properties, differ according to their classes and for which classes these properties differ. In addition, finding differential factors for the use of an alternative method instead of classification according to the main star features is another purpose of the study.. In the study, 2275 exoplanets in F, G, K, M classes and the star class belonging to these exoplanets, their orbital period, transit epoch, impact parameter, transit duration (hour), transit depth (million diameter), planetary radius (planet radius / earth radius), equilibrium temperature (kelvin), insolation flux (earth flux), transit signal-to-noise, stellar effective temperature, stellar surface gravity, stellar radius (stellar radius / earth radius), right ascention, declination and kepler band values were used. The Kruskal-Wallis analysis, was used to find distinctive features in the study, it has been determined that there is a statistically significant difference in at least one of the star classes for the variables of transit epoch, transit duration (hour), transit depth (millionth diameter), planetary radius (earth radius), equilibrium temperature (kelvin), insolation flux (earth flux), transit signal-to-noise, stellar effective temperature, stellar surface gravity, stellar radius (earth radius) and kepler band. A statistically significant difference was found among all class groups in the variables of transit depth, equilibrium temperature, insolation flow, stellar temperature, stellar surface gravity and stellar radius. The specified variables can be used as distinguishing features in the classification of exoplanets.

Keywords: Classification, Exoplanets, Kepler, Kruskal Wallis
JEL-Classification: C10, C30, C89

SINIFLARINA GÖRE ÖTE GEZEGENLERİN FARKLILIKLARININ KRUSKAL WALLİS YÖNTEMİ İLE BELİRLENMESİ

ÖZET

Kepler uzay aracı sayesinde 2600 adetten fazla öte gezegen onaylanmıştır. Onaylanan öte gezegenler ana yıldızlarına göre sınıflandırılmaktadır. Bu çalışmada ana yıldız özelliklerine göre sınıflandırılan öte gezegenlerin sınıflarına göre hangi özelliklerinin değiştiği ve hangi sınıflar için bu özelliklerin farklılık gösterdiğinin tespit edilmesi amaçlanmıştır. Ayrıca ana yıldız özelliklerine göre sınıflandırma yerine alternatif bir yöntemin kullanılması için ayrıcı faktörlerin bulunması da çalışmanın diğer bir amacıdır. Araştırmada NASA'nın oluşturmuş ve daha sonra bilim insanları tarafından düzenlemiş olan "The NASA Exoplanet Archive: Data and Tools for Exoplanet Research" adlı veri seti kullanılmıştır. Çalışmanda F, G, K, M sınıflarında bulunan 2275 öte gezegen ve bu öte gezegenlere ait yıldız sınıfı, yörünge dönemi, geçiş dönemi, etki parametresi, transit süresi (saat), transit derinlik (milyonda çap), gezegen yarıçapı (gezegen yarıçapı/dünya yarıçapı), denge sıcaklığı (kelvin), güneşlenme akışı (dünya akışı), transit sinyalden gürültüye geçiş, yıldız etkili sıcaklık, yıldız yüzey yer çekimi, yıldız yarıçapı

(yıldız yarıçapı/dünya yarıçapı), sağ yükseliş, sapma ve kepler bandı değerleri kullanılmıştır. Çalışmada ayırıcı özelliklerin bulunması için ANOVA analizinin parametrik olmayan alternatifi Kruskal-Wallis analizi kullanılmıştır. Yapılan Kruskal Wallis analizinden sonra geçiş dönemi, transit süresi (saat), transit derinlik (milyonda çap), gezegen yarıçapı (dünya yarıçapı), denge sıcaklığı (kelvin), güneşlenme akışı (dünya akışı), transit sinyalden gürültüye geçiş, yıldız etkili sıcaklık, yıldız yüzey yer çekimi, yıldız yarıçapı (dünya yarıçapı) ve kepler bandı değişkenleri için yıldız sınıflarından en az birinde istatistiksel olarak anlamlı bir farklılığın bulunduğu tespit edilmiştir.Transit derinliği, denge sıcaklığı, güneşlenme akışı, yıldız etkili sıcaklık, yıldız yüzey yer çekimi ve yıldız yarıçapı değişkenlerinde tüm sınıf grupları arasında istatistiksel olarak anlamlı bir farklılık tespit edilmiştir. Belirtilen değişkenler öte gezegenlerin sınıflandırılmasında ayırıcı özellikler olarak kullanılabilirler
Anahtar Kelimeler: Cüce Gezegenler, Kepler, Kruskal Wallis, Sınıflandırma
JEL-Sınıflaması: C10, C30, C89

1. Introduction

The discoveries of exoplanets in the last twenty years have revealed planetary bodies of various sizes and masses around other stars(Rowe et al., 2014). More specifically, the position of these exoplanets around their host stars has a significant impact not only on their detectability probabilities, but also on the atmospheric, chemical composition and capacity to characterize these atmospheres. Moreover, a relatively large example of about 3400 confirmed planets and 4700 other planet candidates allowed us to calculate exoplanet formation rates in our galaxy (Catanzarite & Shao, 2011). NASA's Kepler mission has announced thousands of exoplanet candidates awaiting approval (Borucki et al., 2011). NASA's TESS mission is expected to discover thousands of exoplanet candidates that are earth-size or larger. The classification of exoplanets is limited to nearby stars. Tinetti et al. (2013) in the article they compiled "Are these numbers large enough to provide a meaningful classification of planets as with stars?" (2013) they asked the question. Mass, radius, temperature and chemical compositions of a planetary body are related parameters and interact with the main star(Tinetti et al., 2013).

In this study, based on the ideas of Tinetti et al., it is to determine which planetary properties change according to the classes of the exoplanets that are classified according to the properties of the star in which orbital motion and for which classes these properties differ. In addition, determining the factors that can be used in a classification method according to the planetary characteristics of the alternative exoplanet instead of classification according to star features is another purpose of the study.

2. Literature Summary

The Kruskal Wallis method has been used in many different areas in the literature, but it has not been used in another study in the field of astronomy. In this study, the literature review is limited to the studies on exoplanets in the field of astronomy.

In their study, Saha et al. (2015) classified whether the exoplanets are habitable planets or not by using the distance from the parent star, stellar variability, diversity, kinematic and spectral classification variables.

In their study , Zeng, Jacobsen, and Sasselov (2017) examined whether the radius space of the exoplanets depends on the parent star types, and determined that the radius spaces of the exoplanets depend on the parent star type.

Bugueno et al. (2018) conducted a study to improve the identification of exoplanets from data obtained from Kepler using controlled machine learning methods.

Mena et al. (2019) conducted a study on how machine learning algorithms can be used to discover exoplanets. They created a controlled machine learning algorithm to determine whether the change in the light curve is evidence of a planet.

3. Exoplanet

The first exoplanet was discovered in 1995, and this discovery led to a changeable mindset. The search for exoplanets and their characterization has quickly become a field of scientific research and is also a field supported by funding authorities. Evidence of a low-mass planet orbiting a star near the center of the galaxy at a distance of about 30,000 light-years has been provided, and the first verified micro-lensing planet was reported in 2004.(Perryman, 2018)

Before 1995, the number of planets known was nine, 18 years later Pluto was removed from the planets around the sun, but a thousand planets orbiting other stars were detected, this number is far from final. Current statistical estimates show that each star in the galaxy is host to an average of at least one planet. Considering that the number of stars in the milky way is approximately 1011, there are many more exoplanets to discover.(Cassan et al., 2012)

The rapid spread of spectroscopy in the late 19th century resulted in a large number of stellar spectra. Astronomers have faced enormous difficulties, such as trying to understand them. This was similar to problems previously encountered in the biological sciences and was faced by Linnaeus through his system of classifying living organisms. The system used today has entered the literature with a study published in 1912 after the detailed and comprehensive studies of Annie Cannon and her team at the Harvard Observatory. The same classification is still used, although it has changed since then. The spectral classification of the stars is as in Table 1 below.

Table 1. Star Spectral Classes (Spectral Classes, 2020)

Spectral Class	Effective Temperature (Kelvin)	Color	H Balmer Features	Other Features	$M/M_{Güneş}$	$R/R_{Güneş}$	$L/L_{Güneş}$	Main Rank Life
O	28.000 - 50.000	Blue	Weak	Ionized He^+ lines, strong UV persistence	20-60	9-15	90.000 - 800.000	1 - 10 Myr
B	10.000 - 28.000	Blue - White	Medium	Neutral He lines	3-18	3.0 - 8.4	95 - 52.000	11 - 400 Myr
A	7.500 - 10.000	White	Strong	Strong H-lines, ionized metal lines	2.0 - 3.0	1.7 - 2.7	8 - 55	400 Myr - 3 Gyr
F	6.000 - 7.500	White - Yellow	Medium	Weakly ionized Ca^+	1.1 - 1.6	1.2 - 1.6	2.0 - 6.5	3 - 7 Gyr
G	4.900 - 6.000	Yellow	Weak	Ionized Ca^+, metal lines	0.85 - 1.1	0.85 - 1.1	0.66 - 1.5	7-15 Gyr
K	3.500 - 4.900	Orange	Skinny	Ca^+, Fe, strong molecules, CH, CN	0.65 - 0.85	0.65 - 0.85	0.10 - 0.42	17 Gyr
M	2.000 - 3.500	Red	Skinny	Molecular lines, for instance TiO, Neutral metals	0.08 - 0.05	0.17 - 0.63	0.001 - 0.08	56 Gyr

4. Method

The ANOVA F test is used to test the ratio of more than two groundmasses, but has some assumptions that are often ignored and often violated when used in real world applications.
These assumptions are based on data in each category.

- come from a normal distribution,
- that the groundmass variances in each category are equal and
- It includes that the data are independent from each other.

If these assumptions are compensate, the ANOVA test is a powerful method for determining whether the categories for a variable come from the same groundmass. The Kruskal-Wallis nonparametric test is used when the groundmass means are compared and it is known that the groundmasses do not have equal variances or the groundmasses are not normal.(Montgomery & Runger, 2014; Ostertagová et al., 2014; Ostertagová & Ostertag, 2013)
Wilcox (1995) , as a result of applying a parametric test to a variable that does not come from a normal distribution, concluded that the deviation from normality has some effect on the Type 1 (1995)error, but found that this effect is minimal when the variances are equal. However, Wilcox stated that the effect on the Type 2error is much more pronounced. Therefore, Wilcox stated in own study that it is appropriate to use non-parametric alternatives in cases where assumptions cannot be provided.
Kruskal-Wallis is a nonparametric statistical test that evaluates the differences between three or more independently sampled groups on a single continuous variable that is not normally distributed.(Kruskal & Wallis, 1952; McKight & Najab, 2010)
The purpose of analysing variance is to decide whether each voluminous k number of independent samples come from the same collection. n_j Therefore, the null hypothesis generally states that each voluminous k number of samples come from the same collection. n_j The assumptions of the H statistic proposed by Kruskal-Wallis are as follows.
Assumptions:

1. Data is compiled from k numbers of randomly selected samples of volumes. $n_1, n_2, \cdots, n_k$
2. Observations are independent both within and between samples.
3. The level of measurement should be the least ranking.
4. The variable of interest is continuous.(Gamgam & Altunkaynak, 2013)

The Kruskal-Wallis test is an extension of the two-group Mann-Whitney U (Wilcoxon Rank) test. Therefore, it is a more generalized form of the Kruskal-Wallis Mann-Whitney U test and a nonparametric version of one-way ANOVA (Analysis of Variance). (McKight & Najab, 2010)
In the Kruskal-Wallis test, at least one of the alternative hypothesis groups is different from the others. Therefore, it does not indicate in which group or groups the difference is. For this, it can be determined which groups are different by taking the groups in pairs and performing the Mann-Whitney U test.(Corder & Foreman, 2011)
The popularity of the Kruskal-Wallis test can be attributed to its usefulness in various disciplines such as engineering and manufacturing applications, medicine, biology, psychology, and education.(Ostertagová et al., 2014)

5. Application

In this section, which constitutes the application part of the research, first of all, the purpose of the research, its groundmass, the descriptive statistics of the groundmass, the findings of the research and the analysis results are presented.

5.1. Purpose of the Study

The purpose of this study is to determine which properties of the exoplanets, which are classified according to the properties of the star in orbital motion, change according to their classes and for which classes these properties differ. Another purpose is to determine the factors that can be used in a classification method according to the characteristics of the alternative planet, rather than classification according to star features.

5.2. Research Groundmass

The groundmass of the research is made up of the exoplanets, which are confirmed and classified from the data obtained by the Kepler spacecraft belonging to NASA. Approximately 8100 exoplanets have been detected by the Kepler spacecraft. Of the 8100 detected exoplanets, about 3700 are confirmed as exoplanets. Of the approximately 3700 confirmed exoplanets, 2350 exoplanets orbitals are classified according to the properties of the star on which it moves. The main mass of the study consists of 2275 exoplanets in classes F, G, K, M. In the study, the entire groundmass was included in the analysis. A sampling process has not been carried out.

5.3. Data

The data set named “The NASA Exoplanet Archive: Data and Tools for Exoplanet Research” was used in the study, which was created by NASA and later edited by scientists.
In the study, star class, orbital period, transition period, impact parameter, transit time (hour), transit depth (millionth diameter), planetary radius (planet radius / earth radius), equilibrium temperature (kelvin), insolation flow (world flow), transit the variables of transition from signal to noise, star effect temperature, stellar surface gravity, stellar radius (stellar radius / earth radius), right ascent, declination and kepler band are used. Table 2 shows the variables used in the study, the structure of the variables and the explanations of the variables.

Table 2. Variables and Their Descriptions

Variable Name	Variable Structure	Explanation
Star Class	Categorical	Star Class (F, G, K, M Category Values)
Orbital Period	Metric	The interval between successive planetary transitions
Transition Period	Metric	The time corresponding to the center of the first transition detected on Barianthian Julian Day (BJD) minus a fixed difference of 2,454,833 lasting days. The difference corresponds to 12:00 on January 1, 2009 UTC.
Impact Parameter	Metric	Prediction of the sky at conjunction between the center of the stellar disk and the center of the planetary disk, normalized to the stellar radius.
Transit Time	Metric	The duration of the observed transitions. The duration is measured from the first contact between the planet and the star to the last contact.
Transit Depth	Metric	The rate of stellar flow lost at minimum in planetary transition.
Planetary Radius	Metric	The radius of the planet. Planetary radius is the product of the ratio of the planetary star radius and the stellar radius.
Heat Balance	Metric	The forecast for the planet's temperature.
Insolation Flow	Metric	Insolation flow is another way of giving equilibrium temperature. It depends on stellar parameters (especially stellar radius and temperature) and the semi-main axis of the planet. It is given in units according to what is measured from the Sun for the Earth.

Transit from signal to noise	Metric	Transition depth normalized by the average uncertainty in flux during transitions.
Star Effective Temperature	Metric	Photospheric temperature of the star.
Star Surface Gravity	Metric	The base 10 logarithm of the gravitational acceleration on the star's surface
Star Radius	Metric	The photospheric radius of the star
Right Ascension	Metric	KIC right clarity
Deviation	Metric	KIC deviation
Kepler Band	Metric	Kepler Band

The variable values of the exoplanets used in the study and their descriptive statistics according to star classes are given in Table 3 and Table 4.

Table 3. Descriptive Statistics For The Variables

Variable Name	Average	Standard Deviation	Minimum value	Maximum Value
Orbital Period	26,6246	52,5383	0,3418	1071,2326
Transition Period	157,1546	42,2063	121.4866	589,7198
Impact Parameter	0,4274	0,3324	0.0000	1,3000
Transit Time	4,2875	2,6820	.4279	24.1080
Transit Depth	1117,2974	2656,8289	12,2000	36912,0000
Planetary Radius	2,8347	3,3077	0,2700	77,7600
Heat Balance	837,4009	381,1889	129,0000	3186,0000
Insolation Flow	332,1021	936,5532	0,0700	24240,8300
Transit from signal to noise	79,9717	222,2972	6,0000	3535,6000
Star Effective Temperature	5471,3112	667,7203	2703,0000	7358,0000
Star Surface Gravity	4,4119	0,2353	2,4100	5,2740
Star Radius	1,0639	0,6429	0,1180	19,5300
Right Ascension	290,9504	4,7518	280,2066	301,7208
Deviation	44,3520	3,5630	36,5774	52,1491
Kepler Band	14,3562	1,2008	8,7170	17,4750

The orbital period of 2275 exoplanets used in the study, transition period, impact parameter, transit time (hour), transit depth (millionth diameter), planet radius (earth radius), equilibrium temperature (kelvin), insolation flow (world flow), transit mean, standard deviation, maximum and minimum values of the variables of transition from signal to noise, star-effect temperature, stellar surface gravity, stellar radius (earth radius), right ascent, deviation and kepler band variables were found as given in Table 3.

Table 4. Descriptive Statistics by Star Classes

		Exoplanet Number	Average	Standard Deviation	Standard Error
Orbital Period	F	484	22,7834	40,2463	1,8294
	G	1111	28,2928	51,4701	1,5442
	K	616	27,1049	62,7664	2,5289
	M	64	22,0919	43,5330	5,4416
	Total	2275	26,6246	52,5383	1,1015
Transition Period	F	484	152,2030	33,3684	1,5167
	G	1111	158,6322	44,4103	1,3324
	K	616	158,6202	44,8547	1,8072

	M	64	154,8433	33,4554	4,1819
	Total	2275	157,1546	42,2063	0,8849
Impact Parameter	F	484	0,4649	0,3412	0,0155
	G	1111	0,4186	0,3365	0,0101
	K	616	0,4114	0,3146	0,0127
	M	64	4489	0,3432	0,0429
	Total	2275	0,4274	0,3324	0,0070
Transit Time	F	484	4,5816	2,6715	0,1214
	G	1111	4,4619	2,7417	0,0823
	K	616	3,8602	2,4783	0,0999
	M	64	3,1473	2,8652	0,3582
	Total	2275	4,2875	2,6820	0,0562
Transit Depth	F	484	820,7795	2236,5527	101,6615
	G	1111	1199,7981	2747,2319	82,4211
	K	616	1163,9781	2819,2892	113,5924
	M	64	1478,2500	2218,1942	277,2743
	Total	2275	1117,2974	2656,8289	55,7023
Planetary Radius	F	484	3,0003	5,1079	0,2322
	G	1111	3,0127	2,9543	0,0886
	K	616	2,4742	1,9487	0,0785
	M	64	1,9609	1,1016	0,1377
	Total	2275	2,8347	3,3077	0,0693
Heat Balance	F	484	934,8409	379,7196	17,2600
	G	1111	861,0027	383,2929	11,4994
	K	616	745,0341	354,8948	14,2991
	M	64	579,8281	296,3285	37,0411
	Total	2275	837,4009	381,1889	7,9919
Insolation Flow	F	484	421,4051	947,4899	43,0677
	G	1111	363,2832	1092,5544	32,7783
	K	616	229,7771	592,6619	23,8790
	M	64	100,3422	286,3648	35,7956
	Total	2275	332,1021	936,5532	19,6355
Transit from signal to noise	F	484	71,3407	242,0102	11,0005
	G	1111	91,8539	232,1048	6,9635
	K	616	67,9385	196,7379	7,9268
	M	64	54,7953	50,0674	6,2584
	Total	2275	79,9717	222,2972	4,6606
Star Effective Temperature	F	484	5786,0723	581,8300	26,4468
	G	1111	5608,5257	485,8023	14,5748
	K	616	5099,9481	673,9138	27,1528
	M	64	4283,3438	1015,2744	126,9093
	Total	2275	5471,3112	667,7203	13,9992

Star Surface Gravity	F	484	4,3313	0,2455	0,0112
	G	1111	4,3891	0,2022	0,0061
	K	616	4,4885	0,2357	0,0095
	M	64	4,6799	0,2916	0,0364
	Total	2275	4,4119	0,2353	0,0049
Star Radius	F	484	1,2533	1,0509	0,0478
	G	1111	1,0809	0,4181	0,0125
	K	616	0,9245	0,5043	0,0203
	M	64	0,6776	0,5184	0,0648
	Total	2275	1,0639	0,6429	0,0135
Right Ascension	F	484	291,3164	4,7909	0,2178
	G	1111	290,8909	4,5881	0,1377
	K	616	290,7727	4,9520	0,1995
	M	64	290,9247	5,2296	0,6537
	Total	2275	290,9504	4,7518	0,0996
Deviation	F	484	44,4623	3,7239	0,1693
	G	1111	44,2895	3,5683	0,1071
	K	616	44,3013	3,4413	0,1387
	M	64	45,0905	3,3508	0,4189
	Total	2275	44,3520	3,5630	0,0747
Kepler Band	F	484	13,9462	1,3264	0,0603
	G	1111	14,3720	1,1105	0,0333
	K	616	14,5999	1,1373	0,0458
	M	64	14,8391	1,4463	0,1808
	Total	2275	14,3562	1,2008	0,0252

When the orbital period is examined according to classes, the average values of the F and M classes are very close to each other, and the average values of the G and K classes are also very close to each other. When the transition period was examined by classes, the average for Class F was found to be 152, for M Class 154, and for Class G and K to 158. When the impact parameter is analyzed according to classes, all values are very close to each other. When the transit time was examined by classes, the average of the M class was 3.14, the average of the K class was 3.86, and the average of the F and G classes was found to be approximately 4.5. When the Depth of Transit was examined by classes, the average of the F class was 820.77, the average of the G and K classes was approximately 1170 and the average of the M class was found to be 1478.25. When the planetary radii were examined in terms of classes, the average of F and G classes was 3, the planetary radius of K class was 2.47 and the planetary radius of class M was 1.96. When the averages of equilibrium temperature by classes were examined, the average temperature for class F was 934 Kelvin, class G average temperature 861 Kelvin, class K average temperature 746 Kelvin and class M average temperature 579 Kelvin. When the averages of the classes were examined in terms of sunshine angle, the average of F class was 421, G class average was 363, K class average was 229 and M class average was 100. When the mean of transition from transit signal to noise were examined according to the classes, the average of F class was 71.34, G class average was 91.85, K class average was 67.93 and M class average was 54.79. When the averages of the star effect temperature by classes are examined, the average for the F class is 5786 Kelvin, the average for the G class is 5608 Kelvin, the K class average is 5099 Kelvin and the G class average is 4283 Kelvin. When

the average of stellar surface gravity by classes is examined, it is determined that all class averages are close to each other. When the average of the star radius according to the classes is examined, the average of the F class is 1.2, the average of the G class is 1.08, the average of the K class is 0.92, and the average of the M class is 0.67. When the right rising values are examined, it is determined that the average values for all classes are close to each other. When the averages of the deviation value according to the classes are examined, it is determined that the averages of all classes are close to each other. When the averages of Kepler band values according to classes are examined, it is determined that the averages of all classes are close to each other.

It was wanted to test whether the mean differences found according to the classes were statistically significant or not with the ANOVA test. Kruskal Wallis Analysis, which is a nonparametric alternative to ANOVA, was used in the study since it could not provide the assumptions that ANOVA had. Results of the Kruskal Wallis Analysis are given in the findings section.

5.4. Findings

The results of the Kruskal Wallis analysis are given in Table 5. The hypothesis example of the Kruskal Wallis test is given below. The specified hypothesis was created for all variables during the analysis phase, but was not included in the article.

H_0 :*There is No Difference Between Classes for the Specified Variable.*

H_1 :*At Least One Of The Classes Is Different From The Others For The Specified Variable.*

Table 5. Kruskal Wallis Test

	Kruskal-Wallis H	**df**	**Asymp. Sig.**
Orbital Period	3,730	3	0,292
Transition Period	10,251	3	0,017*
Impact Parameter	7,599	3	0,055
Transit Time	77,245	3	0,000*
Transit Depth	102,455	3	0,000*
Planetary Radius	33,451	3	0,000*
Heat Balance	129,998	3	0,000*
Insolation Flow	130,031	3	0,000*
Transit from signal to noise	15,078	3	0,002*
Star Effective Temperature	474,342	3	0,000*
Star Surface Gravity	314,921	3	0,000*
Star Radius	361,235	3	0,000*
Right Ascension	3,555	3	0,314
Deviation	3,543	3	0,315
Kepler Band	102,746	3	0,000*

When Table 5 is examined, the significance value of the orbital period, impact parameter, right ascension and deviation variables is greater than 0.05. For these variables, that hypothesis is accepted. H_0 In other words , there is no statistically significant difference between classes for orbital period, transition period, impact parameter, right ascent, deviation and kepler band variables.

Transition period, transit time (hour), transit depth (diameter per million), planetary radius (earth radius), equilibrium temperature (kelvin), insolation flow (earth flow), transit signal to noise, stellar effect temperature, stellar surface gravity, significance value of stellar radius (world radius) and kepler band variables is less than 0.05. In other words, there is a statistically significant difference in at least one of the stellar classes for stellar radius (earth radius) and kepler band variables in the transition period, transit time (hour), transit depth (diameter per million), planetary radius (earth radius), equilibrium temperature (kelvin), insolation flow (earth flow), transit signal to noise, stellar effect temperature, stellar surface gravity.

The post hoc test was performed for the variables where there was a significant difference, and the classes in which there was a significant difference were determined. Post-Hoc test showing the differences of the variables is given in Table 6. The hypothesis example of the Post-Hoc test is given below. The specified hypothesis was created for all variables and classes during the analysis phase, but was not included in the article.

H_0 :*There is No Difference Between Two Classes for the Specified Variable.*

H_1 :*There Is A Difference Between Two Classes for The Specified Variable.*

Table 6. Post-Hoc Tests

Variables	Classes	Test Statistics	Standard Error	Std. Test Statistics	Adj. Sig.
Transition Period	F-G	-94,511	35,776	-2,642	0,049*
	F – K	-121,273	39,900	-3,039	0,014*
	F-M	-54,124	87,370	-0,619	1,000
	G-K	-26,762	32,998	-0,811	1,000
	G-M	40,387	84,442	0,478	1,000
	K-M	67,150	86,270	0,778	1,000
Transit Time	F-G	59,713	35,776	1,669	0,571
	F – K	252,023	39,900	6,316	0,000*
	F-M	547,635	87,370	6,268	0,000*
	G-K	192,310	32,998	5,828	0,000*
	G-M	487,922	84,442	5,778	0,000*
	K-M	295,612	86,270	3,427	0,004*
Transit Depth	F-G	-220,214	35,776	-6,155	0,000*
	F – K	-355,000	39,900	-8,897	0,000*
	F-M	-606,436	87,370	-6,941	0,000*
	G-K	-134,786	32,998	-4,085	0,000*
	G-M	-386,222	84,442	-4,574	0,000*
	K-M	-251,436	86,270	-2,915	0,021*
Planetary Radius	F-G	-78,869	35,775	-2,205	0,165
	F – K	66,212	39,899	1,659	0,582
	F-M	284,558	87,370	3,257	0,007*
	G-K	145,081	32,998	4,397	0,000*
	G-M	363,427	84,441	4,304	0,000*
	K-M	218,346	86,270	2,531	0,068
Heat Balance	F-G	141,484	35,776	3,955	0,000*
	F – K	367,013	39,900	9,198	0,000*
	F-M	700,995	87,370	8,023	0,000*

	G-K	225,529	32,998	6,835	0,000*
	G-M	559,511	84,442	6,626	0,000*
	K-M	333,982	86,270	3,871	0,001*
Insolation Flow	F-G	141,488	35,776	3,955	0,000*
	F – K	367,001	39,900	9,198	0,000*
	F-M	701,259	87,370	8,026	0,000*
	G-K	225,512	32,998	6,834	0,000*
	G-M	559,770	84,442	6,629	0,000*
	K-M	334,258	86,270	3,875	0,001*
Transit from signal to noise	F-G	-117,778	35,775	-3,292	0,006*
	F – K	-89,794	39,900	-2,250	0,147
	F-M	-254,450	87,370	-2,912	0,022*
	G-K	27,984	32,998	0,848	1,000
	G-M	-136,672	84,442	1,619	0,633
	K-M	-164,656	86,270	-1.909	0,338
Star Effective Temperature	F-G	288,519	35,776	8,065	0,000*
	F – K	774,698	39,900	19,416	0,000*
	F-M	1059,186	87,370	12,123	0,000*
	G-K	486,179	32,998	14,734	0,000*
	G-M	770,668	84,442	9,127	0,000*
	K-M	284,489	86,270	3,298	0,006*
Star Surface Gravity	F-G	-169,005	35,775	-4,724	0,000*
	F – K	-578,169	39,900	-14,491	0,000*
	F-M	-963,735	87,370	-11,031	0,000*
	G-K	-409,164	32,998	-12,400	0,000*
	G-M	-794,730	84,442	-9,412	0,000*
	K-M	-385,566	86,270	-4,469	0,000*
Star Radius	F-G	203,980	35,776	5,702	0,000*
	F – K	640,836	39,900	16,061	0,000*
	F-M	990,425	87,370	11,336	0,000*
	G-K	436,856	32,998	13,239	0,000*
	G-M	786,445	84,442	9,313	0,000*
	K-M	349,589	86,270	4,052	0,000*
Kepler Band	F-G	-218,207	35,776	-6,099	0,000*
	F – K	-371,824	39,900	-9,319	0,000*
	F-M	-544,045	87,370	-6,227	0,000*
	G-K	-153,616	32,998	-4,655	0,000*
	G-M	-325,838	84,442	-3,859	0,001*
	K-M	-172,222	86,270	-1.996	0,275

When Table 6 is examined, significance value of F - G class groups and F - K class groups for the transition period variable is less than 0.05 and has been rejected.H_0 In other words, there is a significant difference between the F class and the K and G classes. For the transit time variable , the significance value of the class groups F - K, F - M, G - K, G - M and K - M was less than 0.05 and was rejected. H_0 For the transit time variable, it differs for all class groups except F - G class group. For the transit depth variable, significance value for all class groups is less than 0.05 and has been rejected. H_0 Transit depth varies for all class groups. For the planetary radius variable, significance value of class groups F - M, G

- K and G - M has been less than 0.05 and has been rejected. H_0 There is a significant difference between planetary radius G class and K and M classes. There is also a significant difference between the F class and the M class. For the equilibrium temperature variable, significance value for all class groups is less than 0.05 and has been rejected. H_0 Equilibrium temperature differs for all class groups. Insolation variable for all classes belonging to groups significance value turned out to smaller than 0.05, and was rejected. H_0 The insolation differs for all class groups. The significance value of the F - G and F - M class groups for the transit signal to noise variable is less than 0.05 and has been rejected. H_0 There is a significant difference between F class and G and M classes in terms of the transition from transit signal to noise. For the temperature variable with star effective, significance value for all class groups is less than 0.05 and has been rejected. H_0 The star effective temperature differs for all class groups. For the stellar surface gravity variable, significance value for all class groups is less than 0.05 and has been rejected. H_0 Star surface gravity differs for all class groups. For the star radius variable, significance value for all class groups is less than 0.05 and has been rejected. H_0 The star radius differs for all class groups. For the Kepler band variable, the significance value of the class groups F - G, F - K, F - M, G - K and G - M was less than 0.05 and was rejected. H_0 For the Kepler band variable, it differs for all class groups except K - M class group.

6. Result

In this study, it is aimed to determine which characteristics of the exoplanets, which are classified according to the properties of the star in orbital motion, change according to their classes and for which classes these properties differ. In addition, the issue of determining the factors that can be used in classifying the exoplanets according to their own characteristics has been investigated.

In the study, the star class belonging to 2275 exoplanets in classes F, G, K, M, orbital period, transition period, impact parameter, transit time (hour), transit depth (millionth diameter), planetary radius (earth radius), equilibrium temperature (kelvin), insolation flow (earth flow), transit signal to noise, stellar effect temperature, stellar surface gravity, stellar radius (earth radius), right ascent, deviation and kepler band data were used.

After Kruskal Wallis analysis, it has been determined that there is a statistically significant difference in at least one of the stellar classes for transition period, transit time (hour), transit depth (millionth diameter), planetary radius (earth radius), equilibrium temperature (kelvin), insolation flow (earth flow), transit signal to noise, stellar effect temperature, stellar surface gravity, stellar radius (earth radius) and kepler band variables.

A statistically significant difference was found among all class groups in the variables of transit depth, equilibrium temperature, insolation, stellar effect temperature, stellar surface gravity and stellar radius. The specified variables can be used as distinguishing features in the classification of exoplanets.

REFERENCES

Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M., Bryson, S. T., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Latham, D. W., Lissauer, J. J., … Still, M. (2011). Characteristics Of Planetary Candidates Observed By *Kepler*. Ii. Analysis Of The First Four Months Of Data. *The Astrophysical Journal*, *736*(1), 19. https://doi.org/10.1088/0004-637X/736/1/19

Bugueno, M., Mena, F., & Araya, M. (2018). Refining Exoplanet Detection Using Supervised Learning and Feature Engineering. *2018 XLIV Latin American Computer Conference (CLEI)*, 278–287. https://doi.org/10.1109/CLEI.2018.00041

Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., Wambsganss, J., Menzies, J., Williams, A., Jørgensen, U. G., Udalski, A., Bennett, D. P., Albrow, M. D., Batista, V., Brillant, S., Caldwell, J. A. R., Cole, A., Coutures, Ch., Cook, K. H., … Wyrzykowski, Ł. (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature*, *481*(7380), 167–169. https://doi.org/10.1038/nature10684

Catanzarite, J., & Shao, M. (2011). The Occurrence Rate Of Earth Analog Planets Orbiting Sun-Like Stars. *The Astrophysical Journal*, *738*(2), 151. https://doi.org/10.1088/0004-637X/738/2/151

Corder, G. W., & Foreman, D. I. (2011). Comparing More than Two Unrelated Samples: The Kruskal-Wallis H-Test. In *Nonparametric Statistics for Non-Statisticians* (pp. 99–121). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118165881.ch6

Gamgam, H., & Altunkaynak, B. (2013). *Parametrik Olmayan Yöntemler: SPSS Uygulamalı* (5th ed.). Seçkin Yayıncılık.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

McKight, P. E., & Najab, J. (2010). Kruskal-Wallis Test. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (p. corpsy0491). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0491

Mena, F. A., Bugueño, M. C., & Araya, M. (2019). Classical Machine Learning Techniques in the Search of Extrasolar Planets. *CLEI Electronic Journal*, *22*(3). https://doi.org/10.19153/cleiej.22.3.3

Montgomery, D. C., & Runger, G. C. (2014). *Applied statistics and probability for engineers* (Sixth edition). John Wiley and Sons, Inc.

Ostertagová, E., & Ostertag, O. (2013). Methodology and application of oneway ANOVA. *American Journal of Mechanical Engineering*, *1*(7), 256–261.

Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials*, *611*, 115–120. https://doi.org/10.4028/www.scientific.net/AMM.611.115

Perryman, M. A. C. (2018). *The exoplanet handbook* (Second edition). New York, NY : Cambridge University Press.

Rowe, J. F., Bryson, S. T., Marcy, G. W., Lissauer, J. J., Jontof-Hutter, D., Mullally, F., Gilliland, R. L., Issacson, H., Ford, E., Howell, S. B., Borucki, W. J., Haas, M., Huber, D., Steffen, J. H., Thompson, S. E., Quintana, E., Barclay, T., Still, M., Fortney, J., … Geary, J. (2014). Validation Of *Kepler* 'S Multiple Planet Candidates. Iii. Light Curve Analysis And Announcement Of Hundreds Of New Multi-Planet Systems. *The Astrophysical Journal*, *784*(1), 45. https://doi.org/10.1088/0004-637X/784/1/45

Saha, S., Agrawal, S., Bora, K., Routh, S., Narasimhamurthy, A., & others. (2015). ASTROMLSKIT: A New Statistical Machine Learning Toolkit: A Platform for Data Analytics in Astronomy. *ArXiv Preprint ArXiv:1504.07865*.

Spectral Classes. (2020, November 17). *Spectral Classes* [Text]. scheme=AGLSTERMS.AglsAgent; corporateName=CSIRO Australia Telescope National Facility; address=PO Box 76 Epping NSW 1710 Australia; contact=+61 2 9372 4100 (phone),+61 2 9372 4310 (fax); jurisdiction=Commonwealth. https://www.atnf.csiro.au/outreach/education/senior/astrophysics/spectral_class.html

Tinetti, G., Encrenaz, T., & Coustenis, A. (2013). Spectroscopy of planetary atmospheres in our Galaxy. *The Astronomy and Astrophysics Review*, *21*(1), 63. https://doi.org/10.1007/s00159-013-0063-6

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, *65*(1), 51–77.

Zeng, L., Jacobsen, S. B., & Sasselov, D. D. (2017). Exoplanet Radius Gap Dependence on Host Star Type. *Research Notes of the AAS*, *1*(1), 32. https://doi.org/10.3847/2515-5172/aa9ed9