Journal of Emerging Investigators

An Analysis on Exoplanets and How They are Affected by Different Factors in Their Star Systems

--Manuscript Draft--

Manuscript Number:	
Full Title:	An Analysis on Exoplanets and How They are Affected by Different Factors in Their Star Systems
Article Type:	Research Article
Manuscript Classifications:	Manuscript Classifications; Other; Physics; Physics
Keywords:	Exoplanet; Astronomy; Kepler; NASA; Star Systems; Stellar Evolution; Protoplanetary Disk
Corresponding Author:	Logan Paul Selph Brunswick County Early College High School Belville, NC - North Carolina UNITED STATES
Corresponding Author's Institution:	Brunswick County Early College High School
Corresponding Author E-Mail:	selphlogan75@gmail.com
Order of Authors:	Logan Paul Selph
Manuscript Region of Origin:	UNITED STATES
Abstract:	Is there a correlation between star size, and the number of planets that star is likely to support? This paper goes in depth to discover any correlation there is between planets and their stars. Analyzing a dataset in the NASA Exoplanet Archive from August 2017 this paper uses a sample size of around 3500 planets and mathematical statistics to determine if these correlations are present in real star systems. Different elements of exoplanetary star systems were analyzed. First an analysis of all planetary systems in the database, and how many planets were in those systems. Then this was broken down into stellar class, first by individual planets regardless of multi-planet systems, then by the number of these multi-planet systems per stellar class. In every case it was found with high confidence that there is a significant correlation between stellar class and number of individual planets supported, as well as with multi-planet systems. Another analysis was conducted on the probability of planets existing in the habitable zone around their star. This analysis found with high confidence that over 5% of all planets in our galaxy are in these habitable zones.
Additional Information:	
Question	**Response**
Author Comments:	I first became interested in this topic at a summer program called Summer Ventures in Science and Mathematics. Here we were enrolled in summer classes at universities, I has a crash course statistics class and the University of North Carolina Charlotte. We had to choose a topic to do our research paper on, so I looked within the topic of space, which is my true passion, and discovered exoplanet research. I found the whole database of every exoplanet ever discovered and saw it as the perfect way to see if there are any correlations, igniting the flame of my love for exoplanets.

An Analysis on Exoplanets and How They are Affected by Different Factors in Their Star Systems

Logan P. Selph

Real World View of Applied Math and Statistics

Dr. John and Desire` Taylor

Summer Ventures in Science and Mathematics

The University of North Carolina at Charlotte

I. Abstract

Is there a correlation between star size, and the number of planets that star is likely to support? This paper goes in depth to discover any correlation there is between planets and their stars. Analyzing a dataset in the NASA Exoplanet Archive from August 2017 this paper uses a sample size of around 3500 planets and mathematical statistics to determine if these correlations are present in real star systems. Different elements of exoplanetary star systems were analyzed. First an analysis of all planetary systems in the database, and how many planets were in those systems. Then this was broken down into stellar class, first by individual planets regardless of multi-planet systems, then by the number of these multi-planet systems per stellar class. In every case it was found with high confidence that there is a significant correlation between stellar class and number of individual planets supported, as well as with multi-planet systems. Another analysis was conducted on the probability of planets existing in the habitable zone around their star. This analysis found with high confidence that over 5% of all planets in our galaxy are in these habitable zones.

II. Introduction

Extrasolar planetary research is the study of planets outside our own solar system. The goal this research sets out to accomplish is simple, to discover clues about why earth came to be, and to continue a search for life beyond our own planet (NASA). Fortunately, recent innovations in technology have unlocked different methods of discovery, such as transit and radial velocity, increasing exoplanet discovery to levels far beyond those in the previous decade (The Planetary Society). With these discoveries patterns can be seen between different factors in planetary systems and star systems in space. Trends between type of star and number of planets around those types of stars being the focus of this paper.

The research conducted in this paper sets out to find and explain patterns between these factors in planetary systems and how they can be used to find more exoplanets in an easier and more effective way. Many of these correlations revolve around where to find multi-planet systems and what to look for to find them, the more planets there are the more data can be recovered in the smallest amount of time, not to mention the fact that having multiple planets in a system raise the likelihood of finding a planet capable of sustaining the right conditions for life.

The studies conducted in this paper all showed that there are trends between star type and having exoplanets, with G-type stars having over 50% of multi-planet systems in our galaxy. This compounded with another study showing that around 6% of all exoplanets exist in a habitable zone around their star can help in knowing what to filter out and how much to search for before finding the data that one wants. With this information and following the right patterns exoplanet research can reach even greater outputs of relevant data, bringing the answers of why earth is here and if there is life out in the cosmos even closer to being understood.

III. Methods and Results

All datasets used in this paper were acquired from the NASA exoplanet archive in August 2017 and consisted of 3502 confirmed and 5017 candidate planets. Since exoplanets are being discovered and confirmed very frequently, numbers used at this date may differ from anyone else trying to recreate these tests later. With that said, the first data analysis conducted set out to find the composition of planetary systems regardless of star type or having habitable planets. The main question being asked, how many planets are typically in a planetary system?

Figure 1 illustrates the number of systems that exist with a certain number of planets, showing that systems containing only one planet are far more numerous than any other multi-planet system. As the number of planets within a system goes up, the number of those types of systems continues to diminish.

Following this a hypothesis test was conducted to determine what percentage single planet systems make up out of all planetary systems. The type of test conducted was a proportion hypothesis test. The hypothesis being that single planet systems make up more than 75% of all planetary systems in our galaxy. The results follow:

Hypothesis

H_o: $p \leq 75\%$

H_a: $p > 75\%$ (claim)

Data

n= 2611 x= 2030 $\hat{p}= \frac{2030}{2611}$ p= .75 q= .25

Hypothesis Test

$$Z^* = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{(\frac{2030}{2611} - .75)}{\sqrt{\frac{(.75)(.25)}{2611}}} = 3.242782596$$

α–level= .01

The final decision in this hypothesis test (fig. 2) was to reject the null hypothesis, ultimately accepting the alternative. This means that there is sufficient evidence at the α–level of .01 to conclude that over 75% of planetary systems in the Milky Way galaxy are single planet systems. This means that for every random draw of four planetary systems, at least three will only have a single planet.

Next was how many multi-planet systems existed among each different type of star. There are 7 different stellar classification types; O, B, A, F, G, K, and M. These classifications are determined by a star's effective temperature, or the temperature calculated by the radiation it emits. O is the largest and hottest star type on the scale, ranging from 30,000-60,000 K (degrees kelvin), they also give off blue light. B is the second hottest at 10,000-30,000 K, shining blue-white light. Next is the A type which burn at 7,500-10,000 K, shining white. F burns from 6,000-7,500 K, shining yellow-white. G, which is our suns stellar type, burns from 5,000-6,000 K, shining yellow. K burns 3,500-5,000 K and shines yellow-orange. Finally, are the "red dwarf" or M type stars burning at less than 3,500 K and shining red (COSMOS). Temperature is directly influenced by size, the bigger the hotter, but being the biggest is not always the best for harboring exoplanets. Being bigger is much worse, big stars typically blow their protoplanetary disks away due to their very large solar flares and intense solar wind, making the creation of planets highly unlikely. Protoplanetary disks are clouds of matter surrounding stars shortly after their inception, they are responsible for the creation of planets. Solar wind is the constant flow of charged particles from the sun, and solar flares are large ejections of these particles in a short burst. Similarly, smaller stars don't have much matter for this disc, making the creation of planets less common. Medium sized stars typically have a good amount of matter for a disk and don't pose as much as a threat to it as larger stars do, making them excellent candidates for planet creation.

Figure three shows the trend between star type and the number of planets associated with each. The largest and smallest tend to have the least number of planets while the ones in the middle have the most.

Figure four illustrates the trend between the number of multi-planet systems, which is a planetary system containing more than one exoplanet, and the type of star that they formed around. Like in figure three, there were very little systems to show for the biggest and smallest star types, however the medium sized stars, or F-K type, had an abundant amount of multi-planet systems. Of every star type, G had more systems than every other star combined, sticking out as a nursery of sorts for exoplanets. The hypothesis test conducted for this data set aimed to conclude that over half of all multi-planet systems belong to a G-type star. The results follow:

Hypothesis

H_o: $p \leq 50\%$

H_a: $p > 50\%$ (claim)

Data

n= 559 x= 327 $\hat{p}=\frac{327}{559}$ p=.5 q=.5

Hypothesis Test

$$Z^*=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}=\frac{(\frac{327}{559})-.5}{\sqrt{\frac{(.5)(.5)}{559}}}= 4.018071877$$

α–level= .01

The final decision for this hypothesis (fig. 5) was to reject the null hypothesis. This means that there is sufficient data at the α-level of .01 to conclude that over 50% of multi-planet systems in our galaxy have a G-type star.

Following this test is an analysis on G-type stars and the relationship between the number of multi-planet systems and single planet systems that they have. For this test two different populations were used, those of single planets and those with multiple planets. The results follow:

<u>Hypothesis</u>

H_o: $p_1 - p_2 \leq 50\%$

H_a: $p_1 - p_2 > 50\%$

<u>Data</u>

$x_1 = 1100$ $x_2 = 327$ n= 1427 p= .5 $\hat{p}_1 = \frac{1100}{1427}$

$\hat{p}_2$ $\frac{327}{1427}$

<u>Hypothesis Test</u>

$$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\left(\frac{1100}{1427} - \frac{327}{1427}\right) - .5}{\sqrt{(.5)(.5)\left(\frac{1}{1427} + \frac{1}{1427}\right)}} = 2.227511614$$

α–level: .05

For this hypothesis (fig. 6), the final decision was that there was enough evidence to reject the null hypothesis. This means that there is sufficient evidence at the α-level of .05 to conclude that, out of G-type star planetary systems, the probability of finding a multi-planet system is less than the probability of finding a single planet system by about 50%. Overall this means that for every five planetary systems with a G-type star, at most there will be one multi-planet system.

Figure seven illustrates the trend between star type and the number of habitable exoplanets. Again, G is holding the largest number of planets.

Finally, an analysis on potential habitable worlds was collected. The dataset used for this was from the Kepler Objects of Interest program rather than the list of confirmed planets. This dataset had more diversity and depth to it while the confirmed planet list had next to nothing. The restriction for the KOI data was that the planets included had to either be confirmed or a candidate, false positives skew data because they were registered as a planet candidate at one point but were later proven to be something else and not a planet. This restriction put the number of planets in the dataset at 4541, 290 of which were in the habitable zone. The habitable zone, in relation to carbon-based life forms, is the ideal distance from the sun in which liquid water can exist. It is also classified as the region where the planets equilibrium temperature is between 180-310 K and has an insolation flux factor between .25 and 2.2. Equilibrium temperature is the theoretical temperature a planet would be if it were only being heated by its parent star. Insulation flux is the amount of energy per unit area that a planet receives from radiation from its star. The test here set to determine what percent of planets are potentially habitable, the results follow:

<u>Hypothesis</u>

H_o: $p = 6\%$

H_a: $p \neq 6\%$

<u>Data</u>

n= 4541 x= 290 $\hat{p} = \frac{290}{4541}$ p= .06 q= .94

Hypothesis Test

$$Z^* = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{(\frac{290}{4541}) - .06}{\sqrt{\frac{(.06)(.94)}{4541}}} = 1.096009418$$

As figure eight shows, the final decision for this test was to fail to reject the null hypothesis. This means that there was not significant evidence at the α-level of .05 to conclude that the percentage of potentially habitable planets is not 6%. This means that out of all the planets in our galaxy, according to the data, 6% of them are in a habitable region around their star.

IV. Discussion

The hypothesis tests and analysis conducted in this paper aimed to prove that there is a common trend between star type and exoplanets in our galaxy. As seen by the stark similarities in figures 3, 4 and 7 alone, this hypothesis is undoubtedly true. The largest and smallest star types have little to no planets, but the stars whose classification lies in the middle have many more. These trends can be observed in figures 3, 4 and 7 as well, moving right starting at O-type there is little change until F-type is observed where there is a spike in the number of planets, then moving to G-type where the number of planets peaks, then K-type which have numbers mirroring those of F-types, and M-types which drop back down again to levels slightly above those of the largest planets. The most significant hypothesis test, however, was the figure five test analyzing which star type had the most exoplanets. Showing that more than half of multi-planet systems contain a G-type star means that these stars are amazing candidates when searching for exoplanets.

Implementing this data can be done in one main way. This way is to use this information and create a filter to only search within a certain spectrum range to ensure that the highest output of exoplanet candidate systems, in this case systems including G-type stars, are found. As this method could result in a high output of exoplanet discoveries, it would also neglect other types of stars and systems in the galaxy. In a sense, it would be a waste of good information and a deeper understanding of space. This method can be useful so long as it is not used exclusively, other systems out in space need attention too.

In statistics there is always one major concern, biased, skewed, or inaccurate data. This is no exception. A major example of inaccurate data causing a big problem happened back in 2015 when a team of astronomers discovered that around 54% of the “planets” the Kepler space telescope had discovered were not really planets after all, they were false positives (Writer). This means that the telescope thought it saw a planet and that “planet” was researched and catalogued, but, it was either a star or a brown dwarf, which is essentially a small failed star. This is one of many things that could cause problems with data, especially since this is a relatively new field and not as fine-tuned in its methods as other older fields in space. As this data does have a chance of being inaccurate in some places, the only test having a chance of being affected is the on figure 8, about what percent of exoplanets are in the habitable zone for carbon-based life. The only reason this data can be skewed is because the Kepler Objects of Interest data involving confirmed planets and candidates was used instead of the list of all confirmed exoplanets. This was used because KOI (Kepler Object of Interest) has more data and planets, 290 planets that is, that fit into that habitable zone while the other dataset had only 8. Regardless the information in the KOI program has been sorted out since 2015, and the probability that enough planets in this sample turn out to be false positives that it affects my hypothesis test result is slim to none.

As time goes on and technology evolves the data about exoplanets will change and evolve with it. This will either bring about changes in what scientists believed or reinforce what they already do. The very difficult thing about space research is how unpredictable the universe can be, anything can happen. Only time will tell for sure what is correct and what is not.

V. References

[Exoplanet System]. (n.d.). Retrieved August 3, 2017, from http://www.cosmosup.com/wp-content/uploads/2015/07/Astronomers-have-Discovered-A-Truly-Exotic-Star-System-Of-Five-Stars-min1-700x315.png

Exoplanet Exploration Mission Directive. (n.d.). Retrieved August 04, 2017, from https://science.nasa.gov/about-us/smd-programs/ExEP

Gray, R. O., Corbally, C. J., & Burgasser, A. J. (2009). *Stellar spectral classification*. Princeton, NJ: Princeton University Press.

Harvard Spectral Classification | COSMOS. (n.d.). Retrieved August 04, 2017, from http://astronomy.swin.edu.au/cosmos/h/harvard spectral classification

The Planetary Society. How to Search for Exoplanets. (n.d.). Retrieved August 04, 2017, from http://www.planetary.org/explore/space-topics/exoplanets/how-to-search-for-exoplanets.html

NASA Exoplanet Archive [Advertisement]. (n.d.). Retrieved August 1, 2017, from https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=planets

Planetary Systems. (n.d.). Retrieved August 04, 2017, from http://planetary-science.org/planetary-science-3/planetary-astronomy/

Stellar Spectral Types. (n.d.). Retrieved August 04, 2017, from http://hyperphysics.phy-astr.gsu.edu/hbase/Starlog/staspe.html

Summers, M. E., & Trefil, J. (2017). *Exoplanets: diamonds worlds, super-Earths, pulsar planets, and the new search for life beyond our solar system*. Washington, DC: Smithsonian Books.

Writer, M. W. (2015, December 11). Half of Kepler's Giant Exoplanet Candidates Are False Positives: Study. Retrieved August 03, 2017, from https://www.space.com/31320-kepler-giant-exoplanets-false-positives.html

VI. Figure Captions

Figure 1: This figure illustrates how many planetary systems exist based on the number of planets in every system. The common number of planets in a system here is one, but can go as high as seven

Figure 2: This figure illustrates the results of the first hypothesis test with a bell curve, showing that the results were of a high probability of the hypothesis being correct (1% chance incorrect).

Figure 3: This figure illustrates how many planets, regardless of star system, exist among every type of stellar type. Most planets exist among G-Type stars, and the rest trickle down among the other types. O-Types have the least number of planets.

Figure 4: This figure builds off figure 3, comparing stellar class to the number of multi-planet systems analyzed for every star type. Each number recorded for a specific type of star represents a planetary system with more than one planet.

Figure 5: This figure illustrates the results of the second hypothesis test with a bell curve, showing that the results were of a high probability of the hypothesis being correct (1% chance incorrect).

Figure 6: This figure illustrates the results of the third hypothesis test with a bell curve, showing that the results were of a high probability of the hypothesis being correct (5% chance incorrect).

Figure 7: This figure illustrates the number of planets that exist in habitable regions around their star based on different stellar classes. Again, G-Type stars contain the greatest number of habitable region planets, with the rest of the stellar classes decreasing in number as different classes are analyzed.

Figure 8: This figure illustrates the bell curve of the fourth hypothesis test. In this case the curve shows the null hypothesis failing to be rejected, supporting the claim that 6% of planets in our galaxy are in habitable zones with a confidence of 95%.

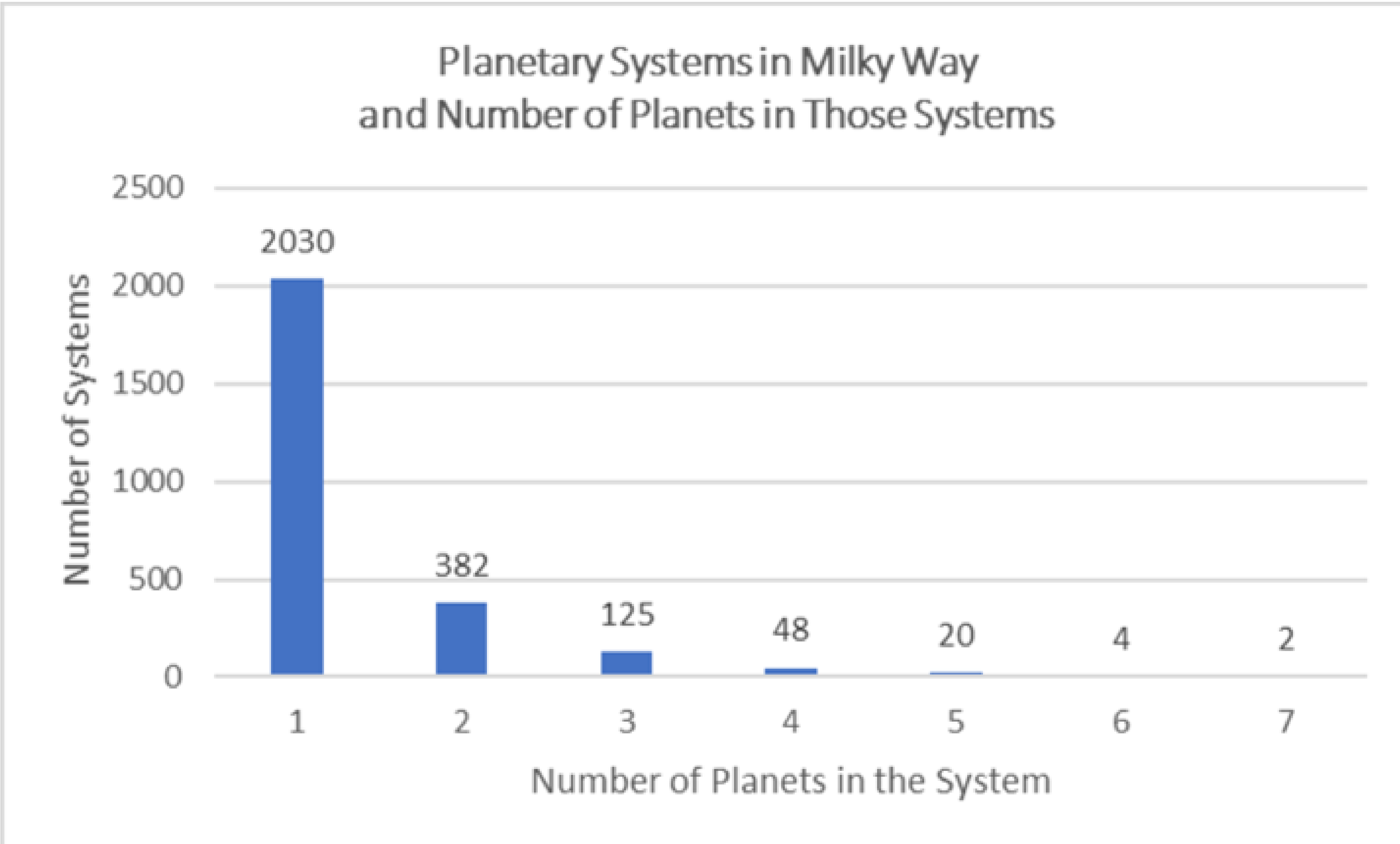

Planetary Systems in Milky Way
and Number of Planets in Those Systems
Number of Systems
2500
2000
1500
1000
500
0
2030
382
125
48
20
4
2
1
2
3
4
5
6
7
Number of Planets in the System

Figure 2

Click here to access/download;Figures;fig. 2.png

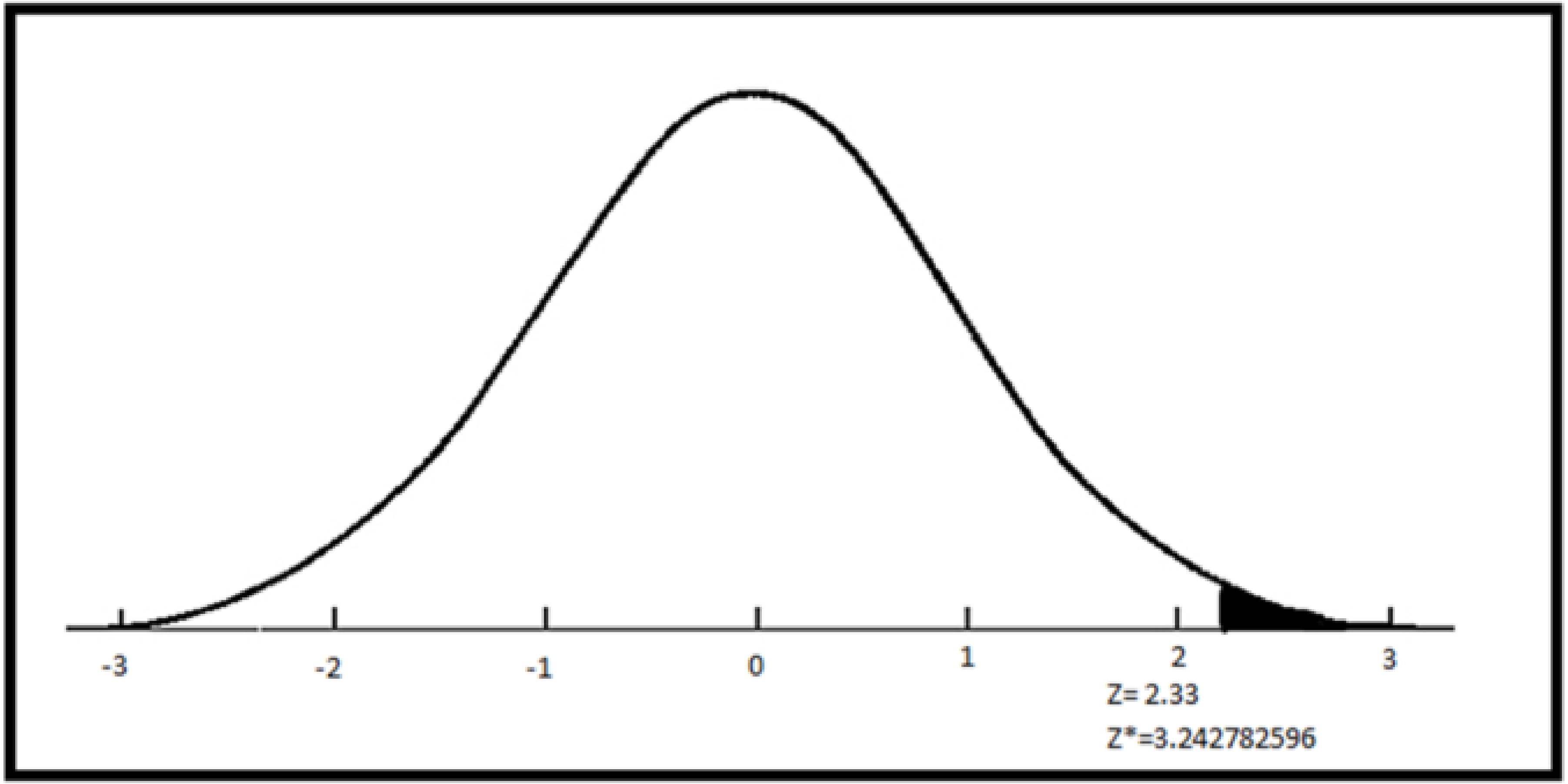
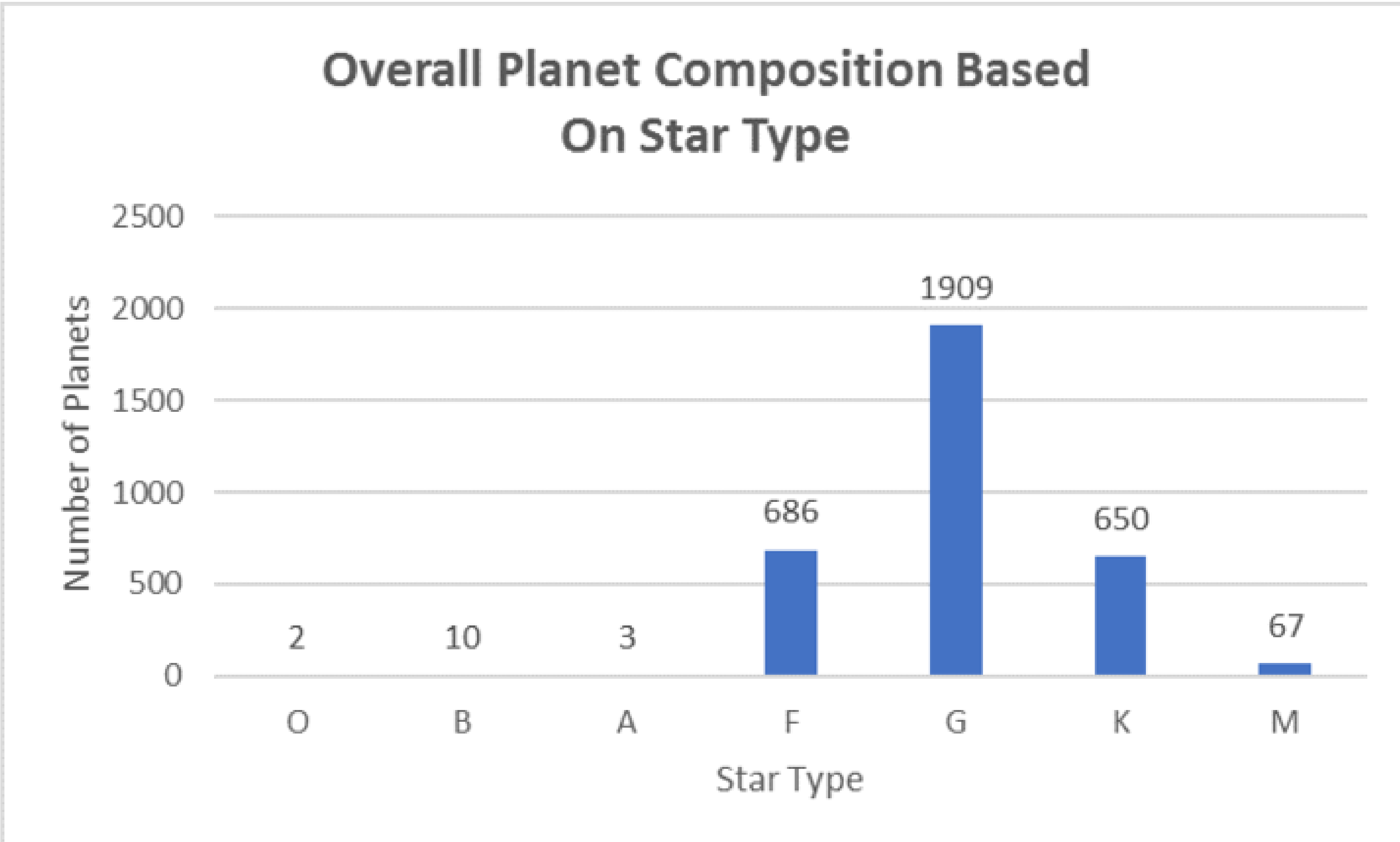


Figure 3

Click here to access/download;Figures;fig. 3.png



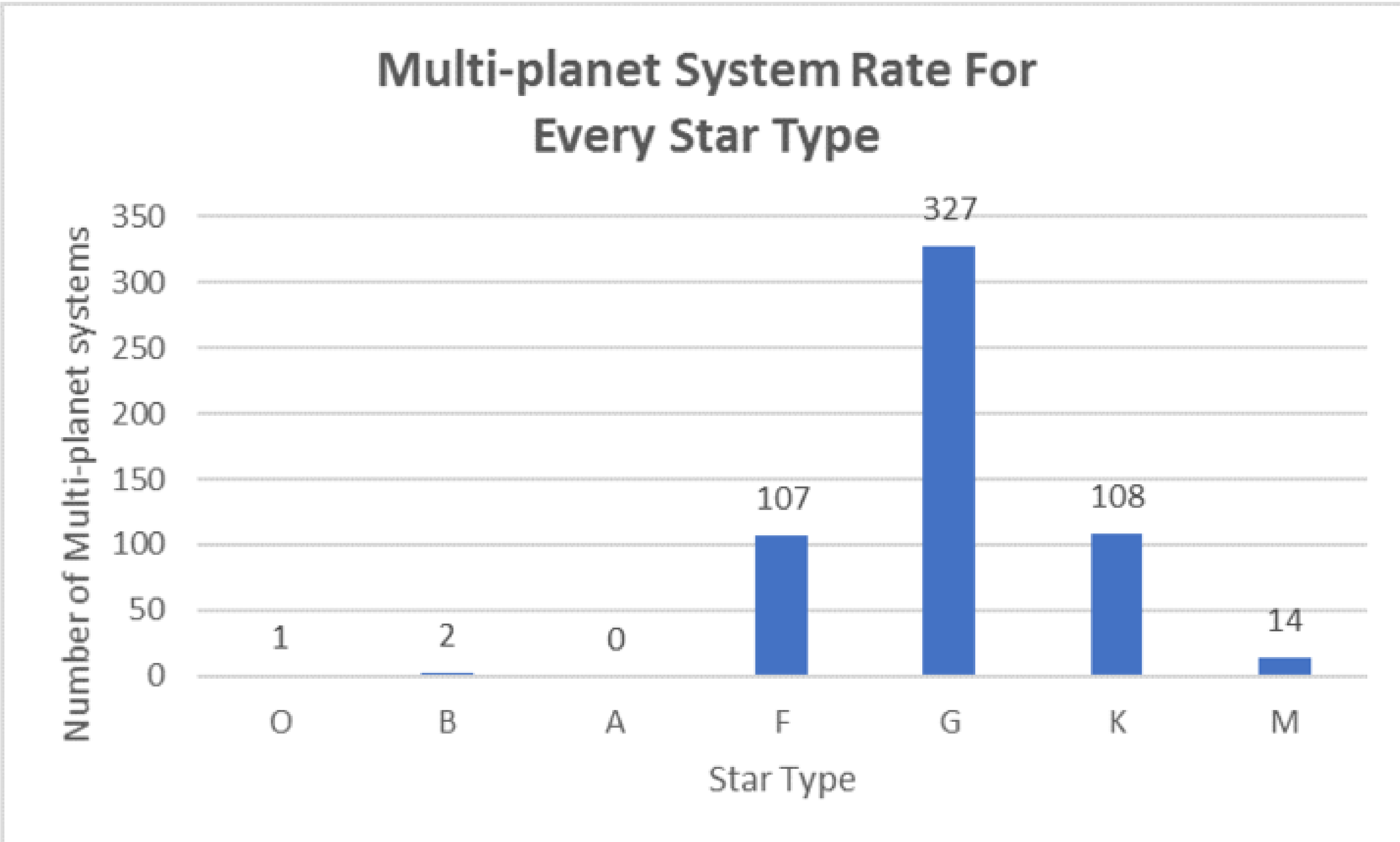

Multi-planet System Rate For Every Star Type
Number of Multi-planet systems
350
300
250
200
150
100
50
0
1
2
0
107
327
108
14
O
B
A
F
G
K
M
Star Type

Figure 5

Click here to access/download;Figures;fig. 5.png

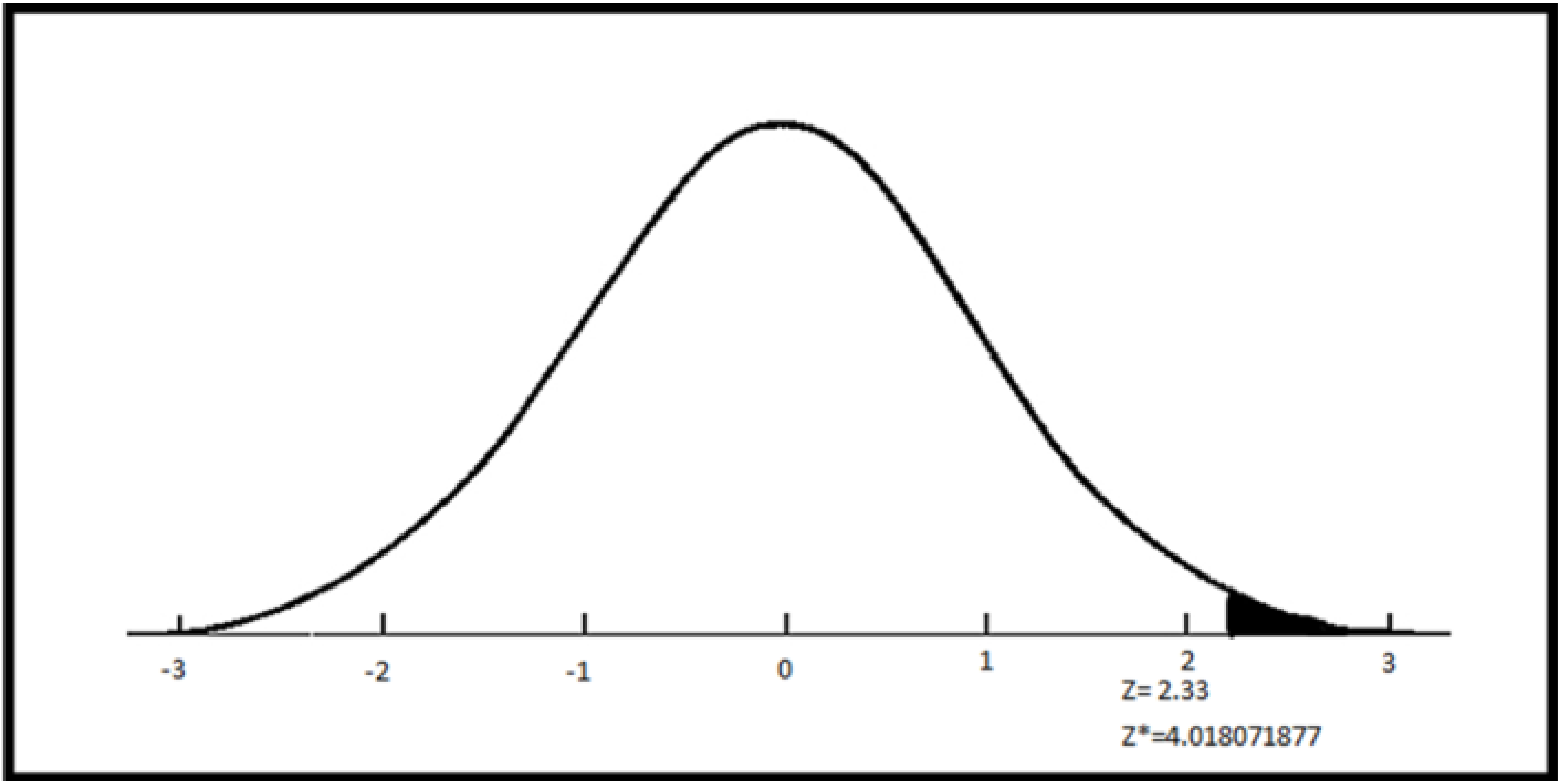


Figure 6

Click here to access/download;Figures;fig. 6.png

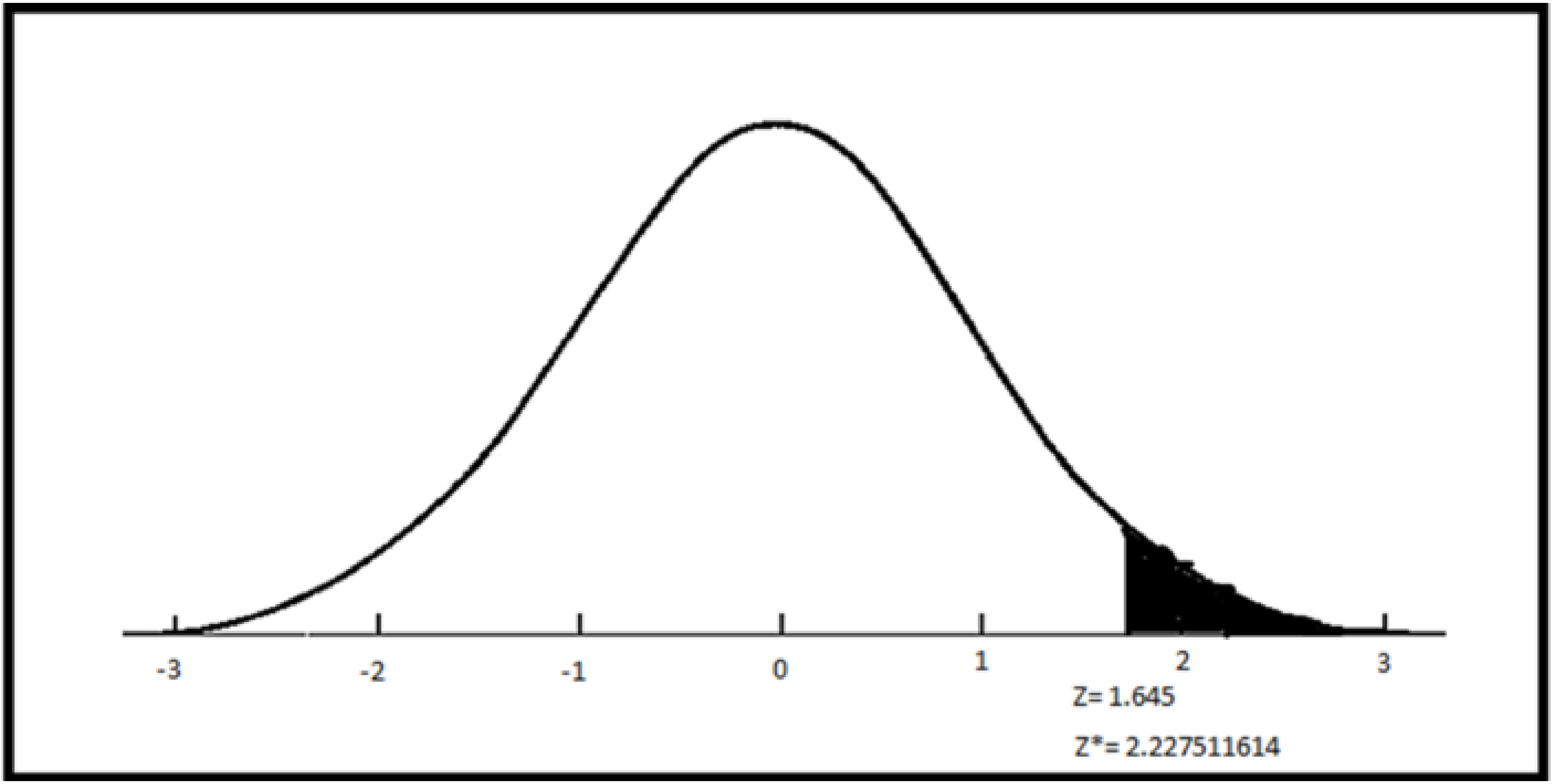


Figure 7

Click here to access/download;Figures;fig. 7.png

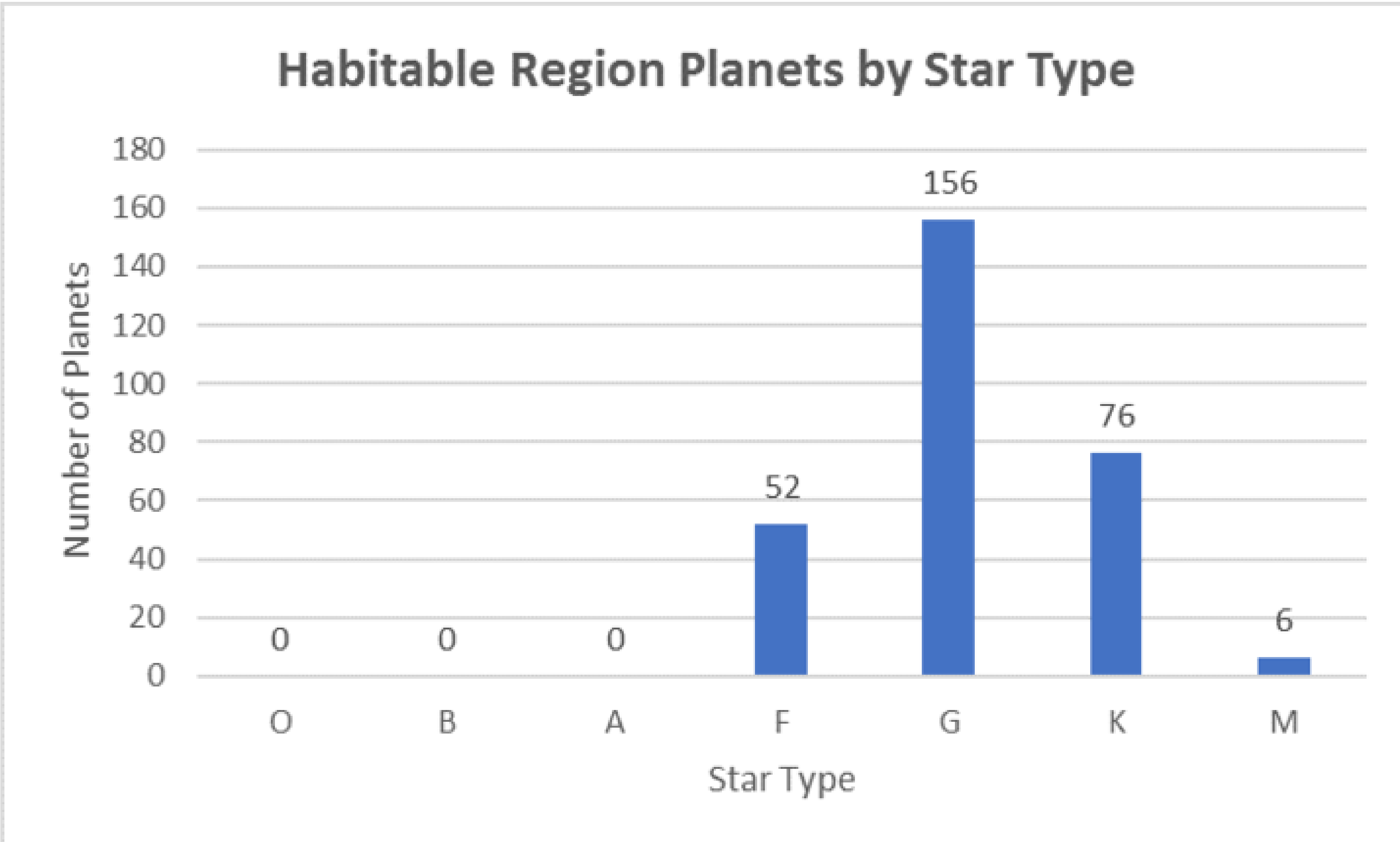


Figure 8

Click here to access/download;Figures;fig. 8.png