

Star Classification and its Connection to Exoplanets

Travis Berger

A World View of Mathematics and Data Analysis

**Dr. John R. Taylor, Mrs. Desiré J. Taylor
and Mrs. Christina L. Turner**

July 17, 2010

Abstract

Is the sun an anomaly in the wide expanses of the universe? Is it unique because of its large collection of planets, including Earth, the seemingly extremely rare life-hosting planet? Or is it just an ordinary star orbited by a bunch of usual, commonplace planets? Only time will tell. However, plenty of data relating a star's spectral class to the existence of extrasolar planets has already been collected, and now this data will be interpreted. Based on the data, a star's spectral classification influences its capability to host planets. More specifically, through hypothesis testing, stars similar to the sun tend to host planets more often than other stars: a conclusion that is both groundbreaking and logical. Also through hypothesis testing, it can be concluded that the majority of stars are without planetary companions.

Background

Exoplanets, or extrasolar planets, are planets found outside the solar system. Technically, the search for exoplanets began about 21 years ago, when astronomers were observing HD 114762, they noticed a slight wobble. This slight wobble resulted from a gravitational pull from some other celestial body, and astronomers concluded that it must be either a brown dwarf or a planet. Although today, astronomers believe that HD 114762 b is a brown dwarf due to more precise measurements (exoplanet.eu).

The first true exoplanet discovered was found orbiting the star 51 Pegasi. Michel Mayor and Didier Queloz, in October 1995, announced their discovery in Florence, Italy at a science conference devoted to cold stars (Mayor & Frei, 2003, p. 1). This was the actual start of the hunt for exoplanets. A few extrasolar planets popped up before 1995, but they were not confirmed until after 1995. For the entire science world, Queloz and Mayor's exoplanet discovery was groundbreaking and revolutionary. From 1995 to present day, plenty of discoveries have

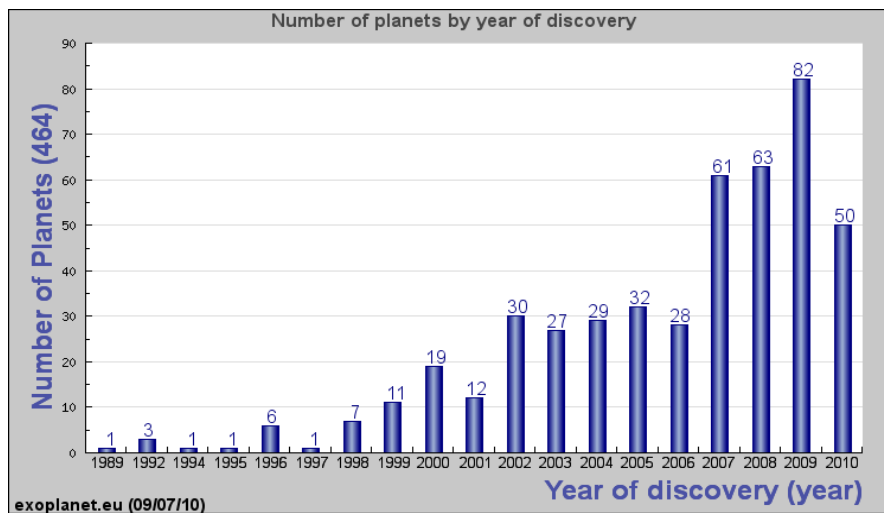


Figure 1: Number of planets by year of discovery

surfaced, as more and more astronomers are joining the search, and more and more effective techniques are employed by astronomers and observatories. As of July 12th, astronomers have

located 464 extrasolar planets (exoplanet.eu).

So, how do astronomers locate these planets? Although the methods may not seem easy or simple, many amateur astronomers could pick up a good telescope, study a particular star, and find the presence of an exoplanet. The three main methods utilized are radial velocity, astrometry, and the transit method. Most planetary detections made so far have exploited the radial velocity technique that focuses on a star's wobble from large-sized planets. As the star wobbles, astronomers can detect a Doppler shift, or a shift in the color of the light emitted. If the star wobbles towards the earth, it emits blue light, and if it wobbles away from the earth, it emits red light (SuperWASP). The star reacts to the planet as the planet orbits around it. With this method, it is possible for one to estimate the planet's minimum mass (ESA). Astrometry, a method similar to radial velocity, measures the position of a star precisely. Then, any wobbling can be detected directly (ESA). Based on the data from exoplanet.eu, 433 planets, 371 planetary systems, and 41 multiple planet systems were discovered using both radial velocity and astrometry.

The final widely used procedure, called transit, employs sensitive instruments to measure the amount of light incoming from a target star. When a planet passes between the star and the earth, the amount of light one receives from the star decreases, so one knows that a celestial body partially eclipsed the target star. Like radial velocity, transit can only identify large planets. So far, 87 planets, 87 planetary systems, and 4 multiple planet systems have been detected by the transit procedure (exoplanet.eu). Other methods – direct imaging, timing, and microlensing – are applied to the search for planets, but they are not as successful due to limited technology. Combined, these three methods have detected only 31 planets. Of the three, direct imaging is the most difficult, but has the most interest because of its revealing nature and visual characteristics.

Direct imaging is the ultimate goal of astronomers, but it will take some time to develop capable technology in telescopes.

Research Question

Having learned how astronomers locate exoplanets and their subsequent methods to do so, can one notice any connections between the stars that are hosts and their planets? Can one notice any type of difference between stars that are hosts and those that are not? More specifically, can one notice a correlation between a star's spectral classification and the presence of exoplanets? In addition, how common are exoplanets: do they form around the majority of stars, or just a few?

This experiment will seek to answer such questions for the curious, the dedicated, and the elite. Using data from solstation.com – a database of stars within 100 light years of the sun and organized by spectral class – this experiment has access to accurate and relevant data that, when analyzed statistically, will reveal correlations between a star's spectral class and the presence of exoplanets, and how common extrasolar planets really are in the expanses of the universe.

Methods

Specific and informative, the data from solstation.com makes great data for statistical analysis. Two branches of statistics exist: descriptive and inferential statistics. Descriptive statistics is the branch of statistics that involves the organization, summarization, and display of data, usually with charts and graphs. Inferential statistics is the branch of statistics that involves using a sample in order to draw a conclusion about a population. When one utilizes both descriptive and inferential statistics, one can accurately portray the characteristics of a sample, and consequently, a population.

To introduce the data itself in the most condensed and numerical form, one should put it into a table. One can see the table's organization by star spectral class and the number of planets by class, along with the total number of stars in each class. The percent with planets was calculated easily by simply dividing the number with planets by the total number of stars in the sample.

Star Spectral Class	B	A	F	G	K	M
# with Planets	0	3	9	34	25	19
# Total	4	76	303	511	948	1000
% with Planets	0%	3.9474%	2.9703%	6.6536%	2.6371%	1.9000%

Table 1: Star Spectral Class

By using descriptive statistics, one can illustrate the data in an easy-to-read, comprehensive graph or chart. As one can see, this pie chart displays the data in an eye-appealing and colorful approach, and it emphasizes the characteristics of the data. This first chart, Exoplanets Found by Star Type, uses the data in the row, # with Planets, and it displays the percentage of the total number of planets that each star spectral class has. For example, A class stars have only 3 planets out of a total of 90 exoplanets, so 3 makes up only 3 % of the data.

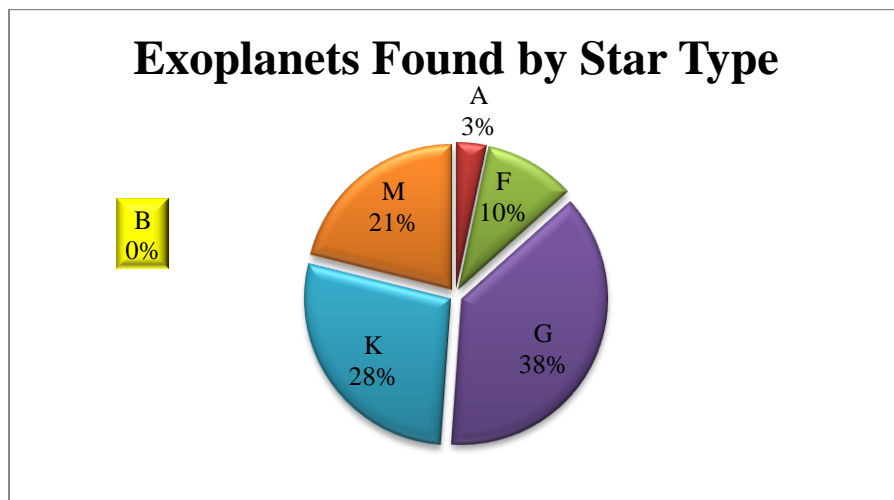


Figure 2.1: Exoplanets Found by Star Type

The second chart, Percent Exoplanets Found by Star Type, presents the data from the row, percentage with Planets, displaying the percentage of the sum of the percentages. Take A spectral class stars again: 3.9474% out of a total of 18% results in A class stars composing 22% of the combined percentages.

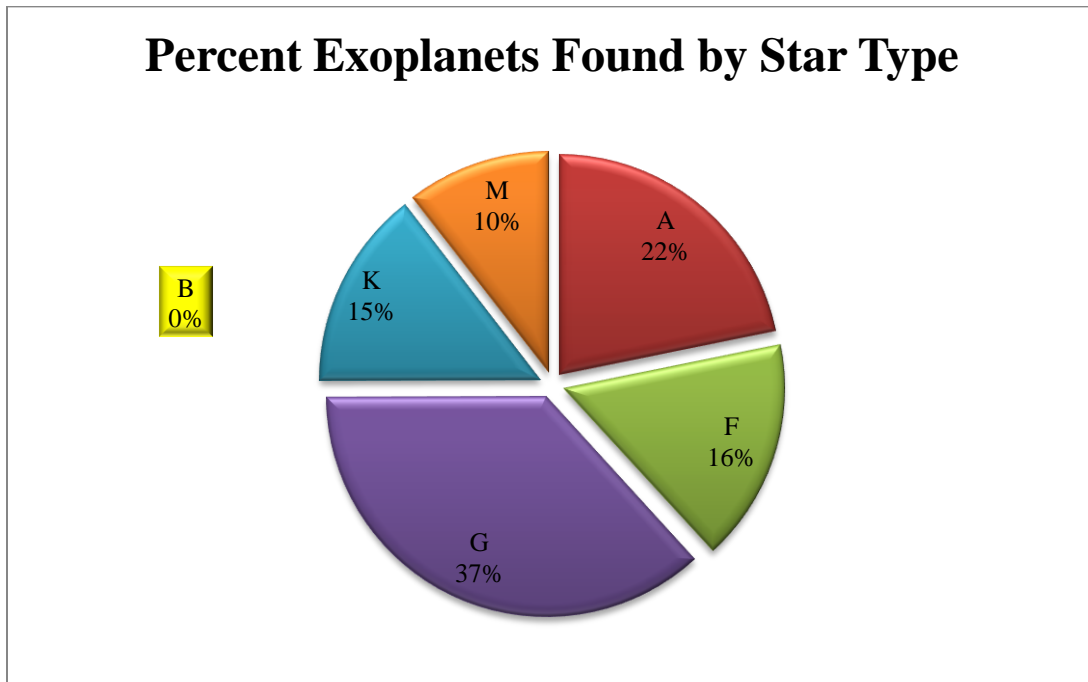


Figure 2.2: Percent Exoplanets Found by Star Type

Inferential statistics may not be as pretty or appealing, but they masterfully reveal very important information about the data, and allow statisticians to draw plenty of useful conclusions. Some inferential methods include: linear regression, probability, sampling distribution, confidence intervals, and hypothesis testing. Based on the raw data, a specific type of test is suited to test the data: hypothesis testing.

Hypothesis testing has two fundamental hypotheses – the null hypothesis (H_0), and the alternative hypothesis (H_a) – of which, the goal is to reject the null hypothesis (H_0). Always stating inequality, the alternative hypothesis (H_a) can be less than, greater than, or not equal to a number or value ($>$, $<$, \neq). Always stating equality, the null hypothesis (H_0) can be at least, at

most, or equal to a specific number or value (\geq , \leq , $=$). In addition, the alternative and the null hypotheses are opposites, i.e.

Because this experiment is trying to prove a star's spectral class affects whether it has a planetary companion or not, the subsequent hypothesis is: Sun-like stars (Type G) will have more than 3% more chance of having an exoplanet than other B, A, F, K, and M classified stars at $\alpha = .10$. The key words: more than 3%, as more than categorizes this hypothesis an alternative hypothesis due to inequality. Divided into two groups by the alternative hypothesis – G class and other class stars – the data will allow one to reject or fail to reject the null hypothesis.

To do a hypothesis test, one must first calculate and designate all of the variables to be determined. Because of the proportional nature of the data, a two-sample proportion hypothesis test must be conducted to figure out whether sun-like stars will have more than a 3% more chance of having a planet when compared to the other classes of stars.

$\alpha = .10$, which means that there is a 10% chance that an error is made.

Sample 1 (Sun-like Stars (Class G))

= number of stars that have planets = 34

= total number of stars = 511

or proportion of sample = .066536

proportion of population

Sample 2 (Other Stars (Classes B, A, F, K, M))

= number of stars that have planets = 56

= total number of stars = 2327

or proportion of sample = .024665

proportion of population

All of the previous variables are used to solve these final three equations:

_____ : The test statistic/value that must fall in the rejection region.

Now that one has all of the equations and variables necessary, hypothesis testing begins.

The first step in a hypothesis test is to state the null and alternative hypotheses:

Hypotheses:

H_0 :

H_a : (claim) (One tail right test)

Critical Values/ Rejection Region: Next, one must calculate the z values to figure out the rejection region, in which the null hypothesis should fall. To calculate the z value, one uses the α level of .10, use the t-chart, look for .10 on One tail, and go to the very bottom where ∞ and .10 intersect. Because this experiment has such a large sample size, it has infinite degrees of freedom, so one copies down the intersected number, which is 1.282.

The area filled to the right is the rejection region: the target.

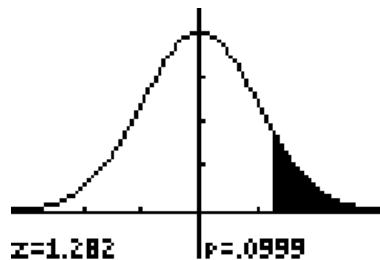


Figure 3.1: Bell Curve Test 1

Test Statistic: The test statistic is all that remains, and all it requires is the plugging in of the values listed above.

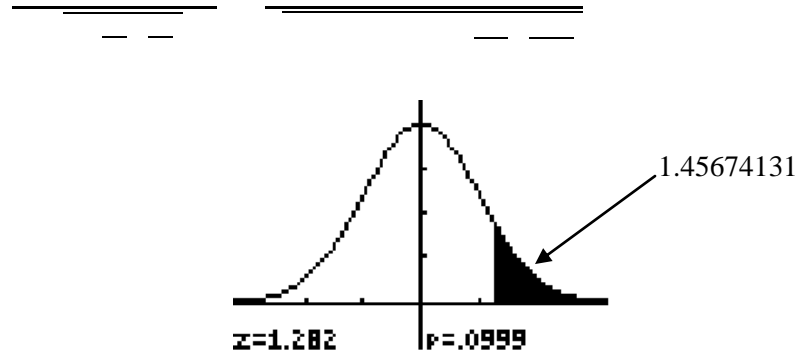


Figure 3.2: Bell Curve Test 1 with Test Statistic

1.456741313 is greater than the critical value (1.282) determined for the rejection region, and because it falls in the rejection region, one rejects the null hypothesis.

Decision: Reject H_0

Conclusion: There is sufficient evidence at the α level of .10 to conclude that the proportion of sun-like stars that have planets is not less than or equal to 3% than other stars, but instead more than 3% than other stars.

The second hypothesis test will test the percent of stars that have a planet orbiting them. Unlike the first time, this trial will not include two separate samples of data, but only one sample. This test will be testing the claim that less than 5 % of stars will be a host to an exoplanet at an α level of .005. The data:

Remember, an α level of .005 indicates a .5% chance of making an error.

= number of stars that have planets = 90

= total number of stars = 2838

or proportion of sample = .031712

proportion of population

$$\frac{\text{---}}{\text{---}} = \text{test statistic} / \text{critical value}$$

State the Hypotheses:

H₀:

H_a: (claim) (Left tail test)

Critical Values / Rejection Region: Using an α level of .005 and a degrees of freedom , one can use the t-chart to figure out the critical value of -2.576, and because it is a left tail test, the rejection region is any value less than -2.576.

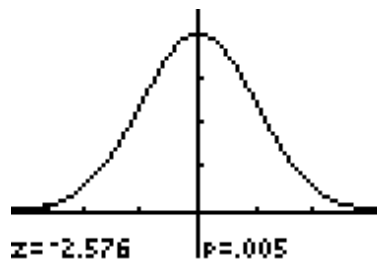


Figure 4.1: Bell Curve Test 2

Test Statistic: To figure out the test statistic or critical value, one plugs the values of the data above into the z equation. Notice that one uses .05 for p:

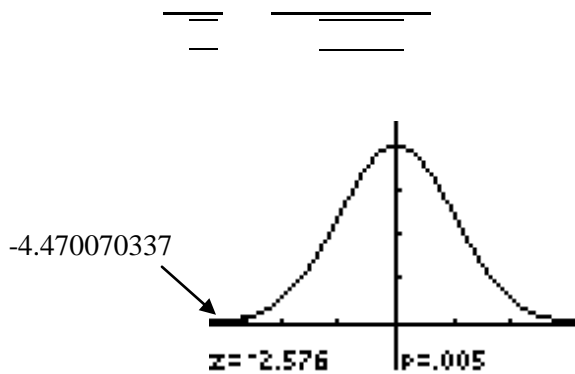


Figure 4.2: Bell Curve Test 2 with Test Statistic

-4.470070337 is less than the critical value of the rejection region, which means that it falls within the rejection region. Therefore, one can reject the H_0 .

Decision: Reject H_0

Conclusion: There is sufficient evidence at the α level of .005 to conclude that the proportion of stars that have exoplanets is not greater than or equal to 5%.

Results

When focusing on the descriptive statistics included in this experiment, one can begin to comprehend a relationship between a star's spectral classification and the presence of an exoplanet. Although the two pie charts may not be the most revealing statistics to the experiment's purpose, they perform perfectly for their own purpose: to show visually a connection between a star's class and whether it has an exoplanet.

The first pie chart, planets found by star type, reveals the percentage of summed planets of each star type. In this example, the pie chart effectively splits the star types into sections of

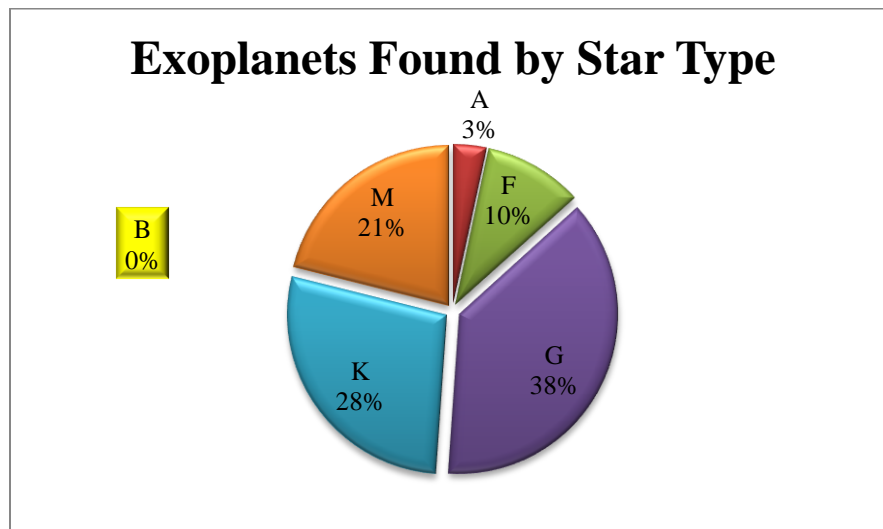


Figure 2.1: Exoplanets Found by Star Type

the pie, so the viewer can see the result: G classified (sun-like) stars have the majority of the exoplanets, at 38%. The second pie chart uses data from the percentage of stars that have planets, so at around 6.6% of a total of around 18%, G stars make up about 37%, again the dominant planet host.

Looking at the inferential statistics, one can conclude even more information from the hypothesis testing. The first hypothesis test dealing with a comparison between different star types allows one to decide “there is sufficient evidence at the α level of .10 to conclude that the proportion of sun-like stars that have planets is not less than or equal to 3% more than other stars.” In other terms, the conclusion exposes that although there is a 10% chance of error; sun-like stars are shown to have more than a 3% increase of exoplanets when compared to other stars. What does this mean? It turns out the sun is not so peculiar when compared to other stars of its type, and if planets will form, sun-like stars are the most likely candidates.

However much it may seem like the sun is not peculiar, the second hypothesis test proves just how rare planets are in this infinite universe. According to the data, with a .5% chance of error, less than 5% of stars are hosts to planets. Maybe the sun is peculiar and unique after all. Although sun-like stars have the greatest chance of forming planets, it is nevertheless very small.

Before the conclusion, there is a specific bias in all of the results reached in this experiment: because current methods of detecting extrasolar planets are still in the development and beginning stages, only certain planets can be detected, if detected at all. As mentioned in the background about the methods that astronomers use to detect exoplanets, most of the methods are only capable of detecting larger-than-Jupiter-size planets. This results in a sampling bias, as for all anyone knows, there could be plenty of smaller, earth-like planets orbiting stars. In

addition, the abundance of stars in the universe and in the Milky Way consequently results in the inability to collect data on all of them, biasing the data even more.

Conclusions and Extensions

The experiment began with two questions: “which spectral class star has the most abundance of exoplanets?” and “how common are exoplanets?” Based on the outcomes from the data, G type (sun-like) stars have the most abundance of exoplanets by more than 3%, and exoplanets are not common, with less than 5% of stars hosting an exoplanet. But what exactly does this mean in the scheme of things, the scheme of the universe? One can deduce that sun-like stars are the best candidates for planetary companions. One can deduce that the sun’s planetary system is a rare find in the large and empty universe. One can deduce that Earth is a truly exceptional gem in the desolate and barren wasteland that is the universe. One can deduce plenty of concepts from the data, but one has to keep in mind the sampling bias. Most likely, astronomers have discovered only a fraction of the planets whose stars astronomers have studied. Because of the sampling bias and the lack of great technology, astronomers overlook many planets, as an estimated “20 to 50%” of stars have at least one planet (activemind.com). Compare that to less than 5%, and one can see the extreme sampling bias of this experiment.

The search for extrasolar planets has always been self-motivated, but an even larger prize occupies the minds of determined astronomers: the detection of life. An application of the planet search has always and will always be the search for life, and maybe even intelligent life. Sir Frank Drake, in 1961, formulated an equation that is still used today that is rightfully known as the Drake Equation (activemind.com). It is supposed to calculate the chances of finding intelligent life somewhere in the Milky Way by using a few variables. The equation is:

where “ N is the number of communicating civilizations in the Milky Way Galaxy, R is the rate at which stars are born each year, f_p is the fraction of stars with orbiting planets, f_{life} is the average number of planets per star in which life can evolve, f_{hab} is the fraction of suitable planets where life develops, f_{int} is the fraction of planets in which intelligent life evolves, f_{com} is the fraction of planetary cultures that are communicating, and L is the average lifetime of these civilizations in years” (Grinspoon, 2003, p. 297). Based on the data from this experiment (using a 5% value for the fraction of stars with orbiting planets and their given values), about 100 communicating civilizations are within the Milky Way Galaxy (activemind.com). Out of at least 100 billion solar systems, only 100 civilizations on the habitable planets are present in the Milky Way, or only .0000001% of stars have earth-like planets with civilizations that can communicate. The Earth is beyond doubt a special planet; a blue, habitable oasis in the black vacuum of space.

The search for planets will continue, and as astronomers’ methods become increasingly accurate and effective, so will the results. From this experiment, one learns that sun-like stars are the best candidates for planets, and planets are relatively rare. In addition, one learns it would be prudent to respect the sampling bias, and try to avoid biasing the data as much as possible. Sadly, the data that this experiment used is not any different from the other data on this topic, so there will be biases whenever collecting data on new and revolutionary subjects such as stars with planets. Again, as the methods evolve and improve, astronomers will detect more stars with planets. The ultimate goal for the present, though, is to find one of those extremely hard to find, earth-like planets that is habitable. When astronomers do, this Exoearth will be considered one of the greatest scientific discoveries ever made, and probably one of the greatest scientific discoveries humans will ever make.

References

Exoplanets. (2010). Retrieved July 12, 2010 from superwasp.org: <http://www.superwasp.org/exoplanets.htm>

File: Fomalhaut with Disk Ring and extrasolar planet b.jpg. (2008). Retrieved July 13, 2010 from Wikipedia.com: http://upload.wikimedia.org/wikipedia/commons/a/a3/fomalhaut_with_disk_ring_and_extrasolar_planet_b.jpg

Frei, P.-Y., & Mayor, M. (2003). *New worlds in the cosmos: The discovery of exoplanets*. Cambridge: Cambridge University Press.

Grinspoon, D. H. (2003). *Lonely planets: The natural philosophy of alien life*. New York: ECCO.

How to find an extrasolar planet. (2007). Retrieved July 12, 2010 from esa.int: http://www.esa.int/esasc/semyzf9yfdd_index_0.html

Keyes, B. (2005). *SETI: The Drake Equation*. Retrieved July 15, 2010 from activemind.com: http://www.activemind.com/mysterious/topics/seti/drake_equation.html

Notable Nearby Stars. (2010). Retrieved July 09, 2010 from Sol Company, Solstation.com: <http://solstation.com/stars.htm>

Number of planets by year of discovery. (2010). Retrieved July 08, 2010 from The Extrasolar Planets Encyclopaedia, exoplanet.eu: <http://exoplanet.eu/catalog-all.php?mdaff=stats#tc>

Schneider, J. (2010). *The Interactive Extra-solar Planets Catalog*. Retrieved July 15, 2010 from The Extrasolar Planets Encyclopaedia, exoplanet.eu: <http://exoplanet.eu/catalog.php>

Appendix

Number of planets by year of discovery:

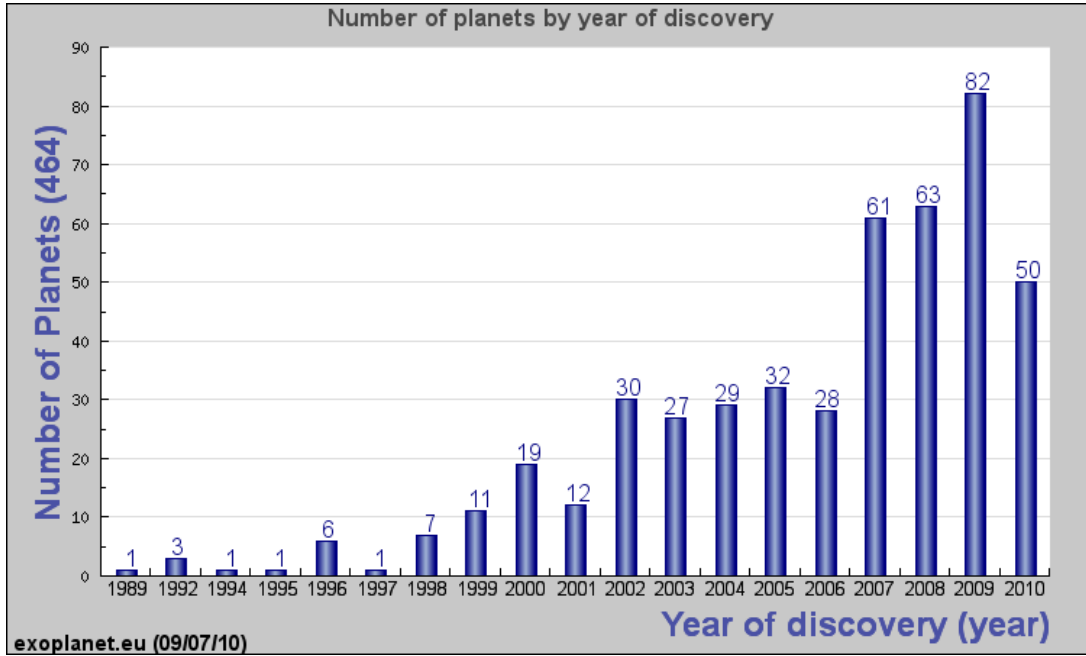


Figure 1

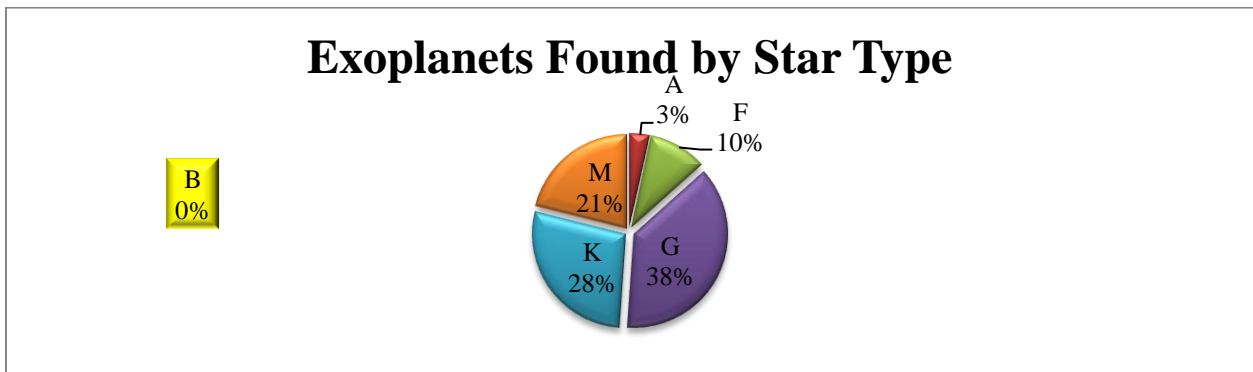
Star Spectral Classification Data Chart:

Star Spectral Class	B	A	F	G	K	M
# with Planets	0	3	9	34	25	19
# Total	4	76	303	511	948	1000
% with Planets	0%	3.9474%	2.9703%	6.6536%	2.6371%	1.9000%

Table 1

Exoplanets Found by Star Type:

Figure 2.1



Percent Exoplanets Found by Star Type:

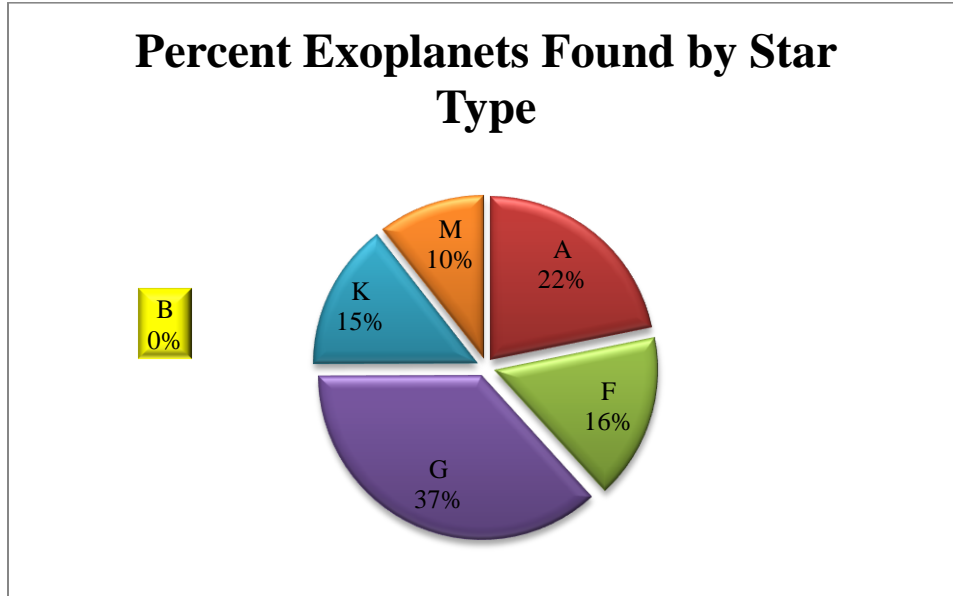


Figure 2.2

Bell Curve Test 1:

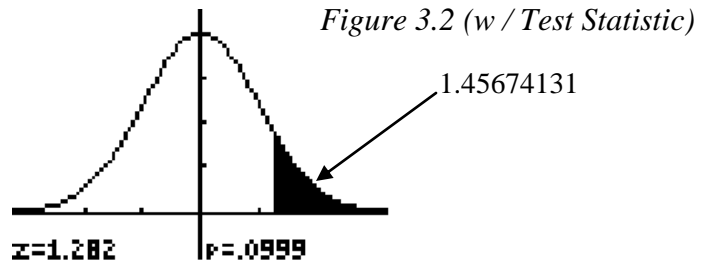


Figure 3.1

Bell Curve Test 2:

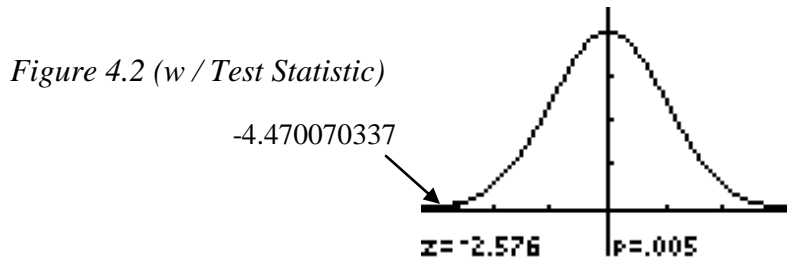


Figure 4.1