Estimating the Hubble Constant and the Curvature of the Universe

Kevin Du

Abstract

This study analyzes observational data from studies regarding distant celestial objects to estimate the Hubble Constant. Using distance, velocity, magnitude, and redshift, two best-fit models were created within Python applying linear regression. Firstly, a linear regression was applied to the entire dataset, but after, the data was separated into 12 chunks, sorted by their distances from the Earth. The results align with modern conclusions of the Hubble Constant, with the calculated value from the linear regression being 50.585 km/s/Mpc.

Introduction

The Hubble Constant (H_0) is a value used to quantify the rate of cosmic expansion, first proposed by Edwin Hubble in 1929 as part of Hubble's Law. According to modern data, the current value of the Hubble Constant is around 69.8 km/s/Mpc, which means that for every megaparsec of distance, an object's velocity increases by 69.8 km/s. Hubble's initial value for the constant was inaccurate due to a lack of technology to precisely measure distance and velocity, but the relationship he found still holds today [1].

Supernovae, stars that are near the end of their lives that explode, emit a large amount of energy, and leave behind a remnant, are central to this study, where the relationship between their distance and velocity directly showed the Hubble Constant. Specifically, for this experiment, type 1A supernovae were used, which involves white dwarfs. They are used to study the Hubble Constant for the limited variability of their brightness and are a type of astronomical object known as standard candles [2]. This means the light emitted is a constant value, and thus, the distance can be calculated from the magnitude value, using the equation relating absolute magnitude (M) and apparent magnitude (m) of an astronomical object [3]:

Equation 1: Distance

$$D_L = 10^{\frac{(m-M)}{5}+1}$$

The critical density is the theoretical density of the universe that would be geometrically flat, defined by:

Equation 2: Critical Density

$$\rho_{crit} = \frac{3H_0^2}{8\pi G}$$

Using the calculated value of the Hubble Constant from the experimental data to get an estimate for the critical density and comparing it to the observed density of the universe allows us to solve for the density parameter: the dimensionless ratio between the critical and observed density.

Equation 3: Density Parameter

$$\Omega_0 = \frac{\rho}{\rho_{crit}}$$

This density parameter can tell us about the future of the universe. If the density parameter is less than 1, then the universe is open and will expand forever. If the density parameter is greater than 1, then the universe is closed and will eventually collapse. Lastly, if the density parameter is equal to 1, then the universe is flat and will expand forever at a decreasing rate [4].

The universe can be one of three shapes: flat, spherical, or hyperbolic. If the universe were flat, it would be like a flat rectangle folded into a cylinder, then a torus. This means that if you were to look far enough ahead, you would see yourself from the back. Light travels across an edge, and it appears in the same place, but from a different perspective. If you were to look in different directions, you would see infinite copies of yourself. However, compared to the two other geometries, the local geometry is the same, meaning we cannot tell for certain what the universe's shape is. Despite this, a flat universe would show patterns, and so far, looking at the Cosmic Microwave Background, none have been discovered. This means that if there were any, they would be outside the observable universe [5].

The second option would be a sphere, but not a conventional two-dimensional sphere. The universe could be a three-dimensional sphere, meaning that it is a set of points that are a fixed distance away from a center point in a four-dimensional space. Each point in a three-dimensional sphere has an opposite point. Unlike a flat universe, spherical universes can be detected locally by measuring cosmic triangles, created by hot and cold spots in the Cosmic Microwave Background. If the angles within these triangles add up to more than 180 degrees, then we live in a spherical universe [6].

Lastly, hyperbolic geometry means the universe opens outward like a saddle, and much faster than flat geometry. This means that a two-dimensional hyperbolic plane cannot fit within normal Euclidean space. In a shape known as the Poincaré disk, the triangles near the edge appear smaller than those at the center, but from a hyperbolic perspective, they would appear the same size. Furthermore, the angles of cosmic triangles in a hyperbolic universe would add up to less than 180 degrees [7]. Out of the three possible geometries, a hyperbolic universe fits least with known information, and a flat universe has become increasingly more likely. It is definitely still possible that we live in a spherical or hyperbolic universe, but it is impossible to tell because humans can only see so far into space; locally, all these geometries are flat [8].

Methods

Hubble analyzed measurements of cepheid variables at Mount Wilson Observatory at the Carnegie Institution of Washington using estimated distances and recession velocities. However, his estimations were inaccurate and resulted in an overestimate of the Hubble Constant due to his primitive technology and early, inaccurate understanding of outer space [9]. Hubble's original data was used again in this experiment, but just like Hubble, the estimation was inaccurate.

Knowing that Hubble's data was flawed, another set of modern data was used in addition to the original. However, Hubble's original data was what was first analyzed and graphed. This dataset contained both distance and velocity values, of which distance was graphed as the x-axis and velocity as the y-axis using the Python libraries pandas, seaborn, and scipy. Although the value calculated for the slope of the line matched what Hubble's value was and showed a clear relationship, the inaccuracy of the data meant that I needed to use more modern and precise data. Thus, the modern data was obtained from the paper "Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples" [10]. The data provided only redshift, magnitude of brightness, stretch, and color values. The distance was calculated from the magnitude value, using equation 1, which relates absolute magnitude (M) and apparent magnitude (m) of an astronomical object.

Furthermore, using the redshift value, recession velocity was found using the relativistic Doppler formula:

Equation 4: Doppler Formula

$$v = c \times \frac{(1+z)^2 - 1}{(1+z)^2 + 1}$$

where $c = 3 \times 10^8$ m/s, the speed of light. Once again, the distance value was plotted as the x-axis and the velocity as the y-axis.

These two values were plotted on a graph, where distance was the independent variable and velocity was on the y-axis, and a linear regression was applied to provide an estimation for the Hubble constant. This was done using the Pandas, Seaborn, Matplotlib, json, SciPy, and NumPy libraries. However, it was found that this estimate was inaccurate, and thus, the modern dataset was divided into 12 different sections, with each bin containing 36 to 37 supernovae, split by distances. A linear regression was applied to each subset individually and the slopes of each were recorded. Since the universe expands at an accelerating rate and the Hubble Constant changes over time, the only relatively accurate estimate for the Hubble Constant would occur within the first subset, as the observations for those objects would be the most recent.

Results

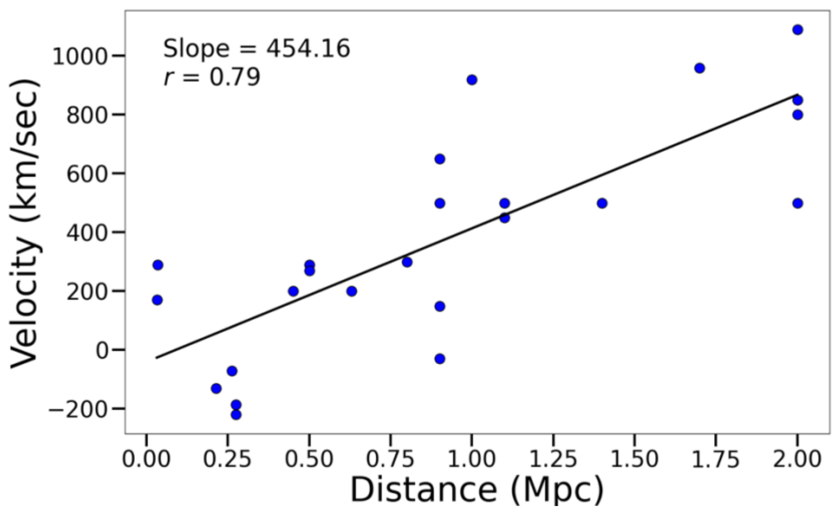

Figure 1. Hubble's original data plotted in a distance vs. velocity graph. The line of best fit is plotted, and the slope is the estimated value of the Hubble constant.

The slope of the linear-fit line in Figure 1 is 454.16 and has an R-value of 0.79, showing a strong correlation between the best-fit line and the data itself. However, due to inaccurate measurements and the lack of modern technology to help do so, his value for the constant itself was about 6.5 times greater than the modern value.

Applying a linear fit to the dataset as a whole does not give an accurate estimate of the Hubble Constant because the data is not linear, but rather, takes the shape of an increasing concave down curve. This occurs because the Hubble Constant has been increasing over time due to the accelerating expansion of the universe.

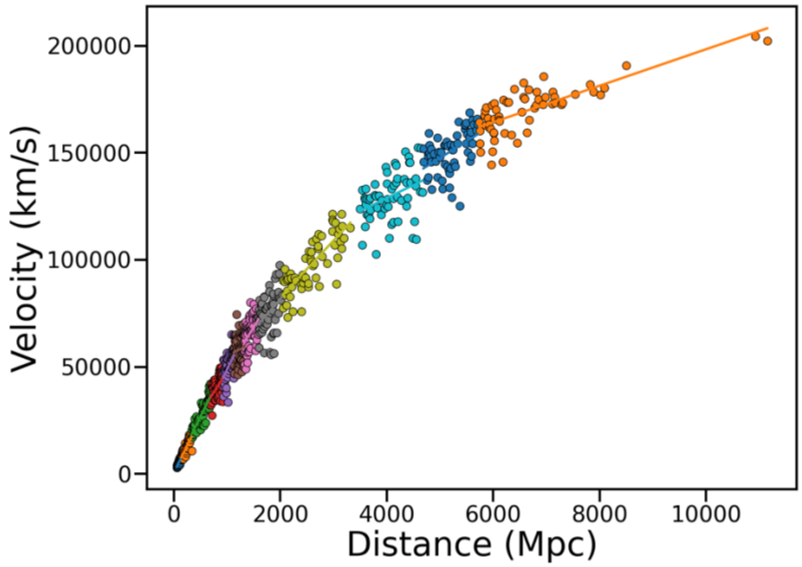

Figure 2. Data from “Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples” by Betoule, M., et. al. Data is grouped into distinctly colored clusters by distance, and a linear fit is applied to each individually.

Figure 2 is plotted using modern data from “Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples” by Betoule, M., et. al. Unlike Figure 1, it utilizes 12 linear regressions applied to 12 individual subsets of the data, instead of a singular linear fit to the entire dataset. Doing so allows for a much better fit for the regression and gives a far more accurate estimation of the Hubble Constant. As stated within the Methods section, only the first cluster of data will give the most accurate estimate, and the slope calculated for that group of data was 50.585 ± 3.460. This value for the Hubble Constant closely aligns with current estimates and shows the effectiveness of clustering data.

Table 1. Dataset from Betoule, M., et. al. divided into 12 groups based on distance with each group’s calculated Hubble Constant value

Distance (megaparsecs):	Hubble Constant Value (km/s/Mpc):
0	50.585
140.496	49.456
342.340	50.168

697.997	39.275
927.660	51.851
1133.754	13.066
1324.727	40.230
1592.880	28.310
2046.350	26.580
3498.685	12.138
4690.101	15.597
5742.035	8.521

Using the first cluster's slope for the Hubble Constant and the critical density equation, the calculated value of critical density is 4.77×10^{-27}. First converting the calculated Hubble Constant value to metric units by multiplying by 1000 and then dividing by 3.1×10^{22}, then squaring it and multiplying by three. Lastly, I divided that number by $8\pi G$ to get the final critical density value. Using this Hubble Constant value and comparing it to the average observed density of the universe, which is a value of 0.85×10^{-26}, the critical density value is smaller than the observed density [11] This means that using just my calculations for equation 3, the density parameter is greater than 1; the range of possible values as calculated from the error is between 1.550 and 2.038, and the median is 1.782:

$$\Omega_0 = \frac{\rho}{\rho_{crit}} = \frac{0.85 \times 10^{-26}}{4.77 \times 10^{-27}} = 1.782 > 1$$

As stated in the introduction, if the density parameter is greater than 1, that means that the universe is closed and will eventually collapse back on itself.

Due to the acceleration of the expansion of the universe, the Hubble Constant has been increasing over time. As is shown in the graph of the data, the slopes of clusters of more distant supernovae is significantly smaller than those closer, showing that the Hubble Constant was once smaller in the past. Looking at the slope of the closest cluster will give the most accurate estimate of the Hubble Constant at the present day. The current estimate of the Hubble Constant is either 67.7 km/s/Mpc or 73 km/s/Mpc, but the slope of the first cluster is around 50.6 km/s/Mpc [12]. While the calculated estimate is around 25% lower than the accepted value, it is decently accurate and much more so than Hubble's initial value.

Conclusion

This project used both historical and modern datasets to estimate the Hubble Constant and explore the curvature of the universe. By analyzing supernovae data—specifically redshift and magnitude—from the Betoule et al. paper, I calculated distances and velocities to determine the Hubble Constant using linear regression models. The initial linear regression across the entire dataset yielded an overly simplistic and inaccurate estimate due to the nonlinear nature of the data. By splitting the data into 12 distance-based clusters and performing regressions on each subset, I obtained a more reliable estimate from the nearest cluster: 50.58 km/s/Mpc. Though lower than the current accepted values (ranging from 67.7 to 73 km/s/Mpc), this estimate is a substantial improvement over Hubble's original overestimate and provides valuable insight into the evolving rate of cosmic expansion.

The figure work in this study demonstrated the importance of segmenting astronomical data to account for temporal and spatial variation in universal expansion. One area for further improvement could be visualizing the change in slope across each of the 12 subsets in a composite graph, revealing how the Hubble Constant varies with distance and time. Additionally, more rigorous statistical analysis—such as calculating error margins or confidence intervals for each cluster's slope—could enhance the reliability of the findings. Future projects could build upon this work by incorporating other standard candles or gravitational wave data, which are emerging as powerful tools for independent Hubble Constant measurements.

Limitations

Error propagation plays a role in the results. My calculated density parameter is between 1.550 and 2.038, with the central value being 1.782, but it is very possible that the range of possible values could include 1, meaning my results would instead align with modern studies showing a flat universe. As of right now, my results show that the universe is closed and will eventually collapse back on itself. Using formulae for propagation of uncertainty and variance, it is possible to find that uncertainty and arrive at an even more accurate answer for the density parameter.

Future Steps

An especially intriguing direction for further study is the "Hubble tension," the current disagreement between values derived from local observations (like supernovae) and those extrapolated from the early universe (like cosmic microwave background measurements). More specifically, when measuring the Hubble Constant using calibrated distance ladder techniques, the value of it is around 73 km/s/Mpc, but when using the Cosmic Microwave Background, the Hubble Constant is closer to 67.7 km/s/Mpc. The Hubble Constant must be one of these two values, but you cannot simply take the average of 70.35 and declare that to be the Hubble Constant. This discrepancy, which remains unresolved, suggests that there may be unknown physics influencing cosmic expansion [13]. Investigating this tension using refined datasets, alternative regression techniques (e.g., Bayesian methods), or by introducing new cosmological parameters could shed light on one of modern cosmology's most compelling mysteries. A new observatory–the Rubin Observatory– in Chile is in the process of mapping out the entire visible Southern Hemisphere sky and create a ten-year time lapse using photos taken every day to help scientists better understand dark matter, dark energy, the Solar System, the Milky Way, and other interstellar objects [14]. The data that comes from the Rubin Observatory could very well help scientists clear up the Hubble Tension and arrive at a final answer for the value of the Hubble Constant.

Acknowledgements

I would like to thank my mentors Husni Almoubayyed and Jacob Wisser for offering their guidance and wisdom to me throughout my work on this project.

References

[1] W. L. Freedman and B. F. Madore, “The Hubble Constant,” *Annu. Rev. Astron. Astrophys.*, vol. 48, no. Volume 48, 2010, pp. 673–710, Sept. 2010, DOI: 10.1146/annurev-astro-082708-101829.

[2] B. Leibundgut, “Are Type Ia Supernovae Standard Candles?,” *Astrophys. Space Sci.*, vol. 290, no. 1, pp. 29–41, Feb. 2004, DOI 10.1023/B:ASTR.0000022162.13380.39.

[3] “6.3: Standard Candle - Physics LibreTexts.” [Online]. Available: https://phys.libretexts.org/Courses/Chicago_State_University/PH_S_1150%3A_Basic_Astronomy/06%3A_Measuring_Cosmic_Distances/6.03%3A_Standard_Candle

[4] “Density Parameter | COSMOS.” [Online]. Available: https://astronomy.swin.edu.au/cosmos/*/Density+Parameter

[5] “What Is the Geometry of the Universe? | Quanta Magazine.” [Online]. Available: https://www.quantamagazine.org/what-is-the-geometry-of-the-universe-20200316/

[6] Z. Yu, “The Flatness of The Universe Through Theory and Observations,” *Highlights Sci. Eng. Technol.*, vol. 112, pp. 314–318, Aug. 2024, DOI 10.54097/b816cb51.

[7] M. P. Hitchman, *Geometry with an Introduction to Cosmic Topology*. in Open textbook library. Department of Mathematics, McMinnville, Oregon: Michael P. Hitchman, 2020.

[8] “What shape is the universe?” [Online]. Available: https://www.astronomy.com/science/what-shape-is-the-universe/

[9] “A RELATION BETWEEN DISTANCE AND RADIAL VELOCITY AMONG EXTRA-GALACTIC NEBULAE by Edwin Hubble.” [Online]. Available: https://apod.nasa.gov/debate/1996/hub_1929.html

[10] M. Betoule *et al.*, “Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples,” *Astron. Astrophys.*, vol. 568, p. A22, Aug. 2014, DOI: 10.1051/0004-6361/201423413.

[11] P. a. R. Ade *et al.*, “Planck 2013 results. XVI. Cosmological parameters,” *Astron. Astrophys.*, vol. 571, p. A16, Nov. 2014, DOI: 10.1051/0004-6361/201321591.

[12] M. Rameez and S. Sarkar, “Is there really a Hubble tension?,” *Class. Quantum Gravity*, vol. 38, no. 15, p. 154005, July 2021, DOI 10.1088/1361-6382/ac0f39.

[13] “Hubble tension or distance ladder crisis? | Phys. Rev. D.” [Online]. Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.110.123518

[14] “Legacy Survey of Space and Time (LSST) | Rubin Observatory.” [Online]. Available: https://rubinobservatory.org/explore/how-rubin-works/lsst