Henry Hitch

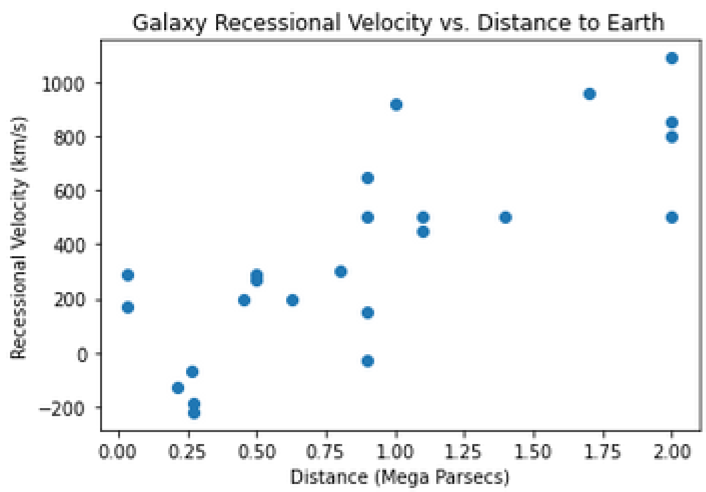
Mark Rast

ASTR 3800

Project 4

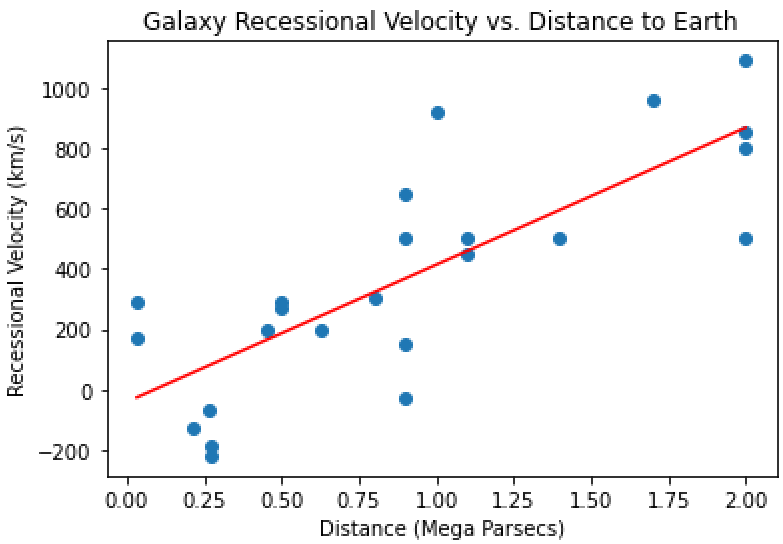
<u>Write Up</u>

After reading in Hubble's original data, I created a scatter plot of the nebulae distance vs. recessional velocity:



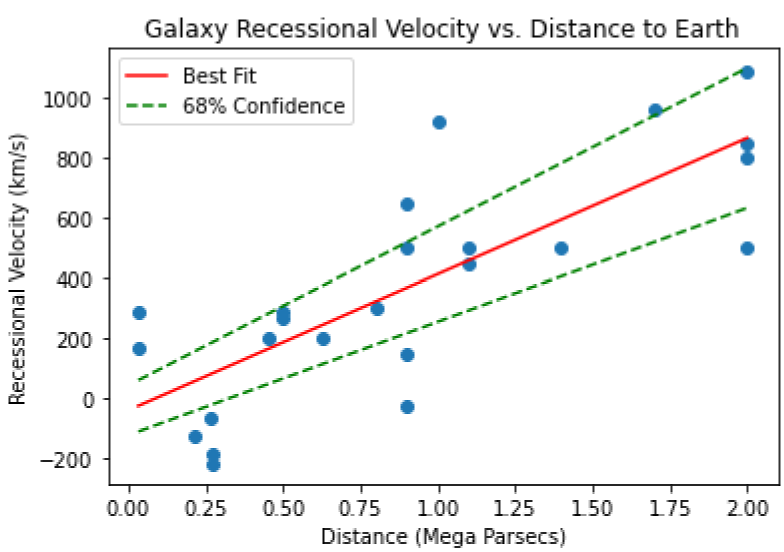
My plot looks very similar to Hubble's original plot, except the axes have slightly different scaling (and I labeled them correctly). Using "chi-by-eye" it's easy to tell there's a positive linear relationship between recessional velocity vs. distance.

Next, I used the Hubble data with np.polyfit() to determine the slope and intercept of the best-fit line to the data, and the uncertainties in these constants. Then, I overplotted the best-fit line over my scatter plot from above:



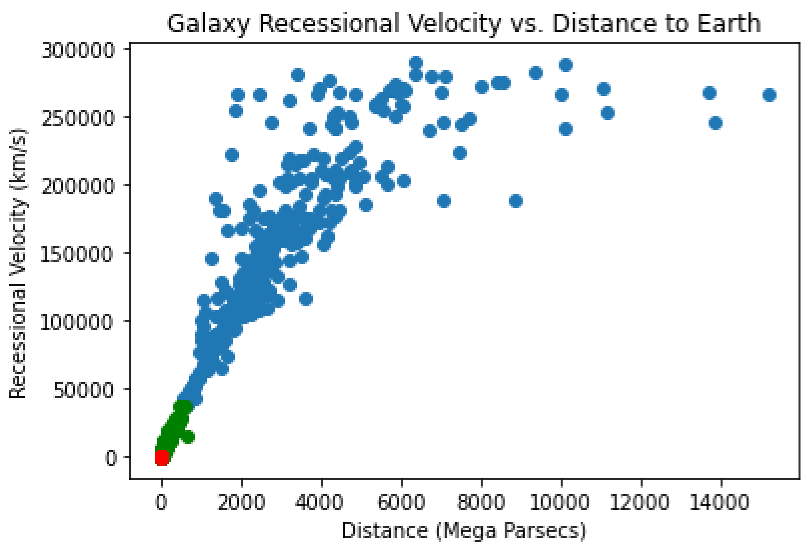
With m = 454 and mVar = 75, this slope is very close to Hubble's value of 465 +/- 50 km/sec/Mpc.

Using the variance in slope and intercept, I overplotted the lines that represent 68% confidence (1 sigma) in the best-fit line:



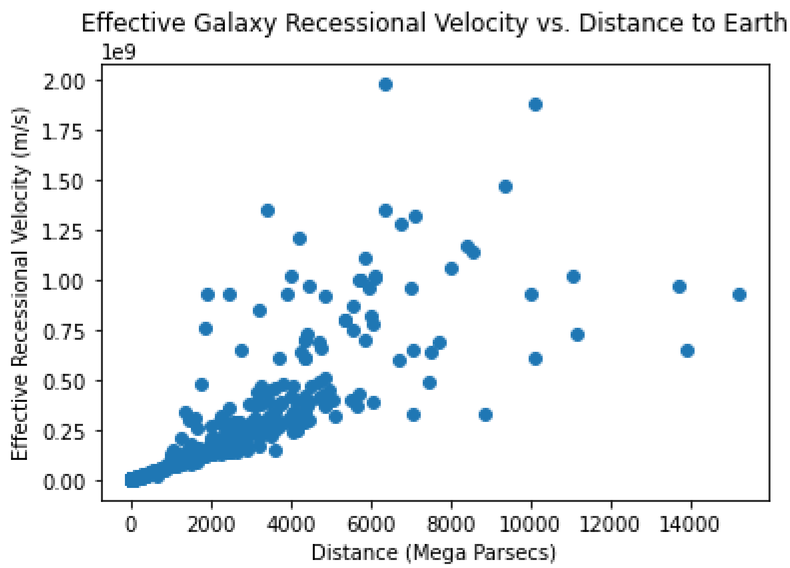
Next, I calculated the Pearson correlation coefficient and p-value between Hubble's two variables, and found that $r = 0.79$ and $p = 4.5e\text{-}6$. Therefore, since r is close to +1 and p is close to 0, this demonstrates a fairly strong positive linear relationship between the velocities and distances.

Then, I extracted more galaxy data from the other csv files, and scatter plotted ALL the velocity vs. distance data on one plot:

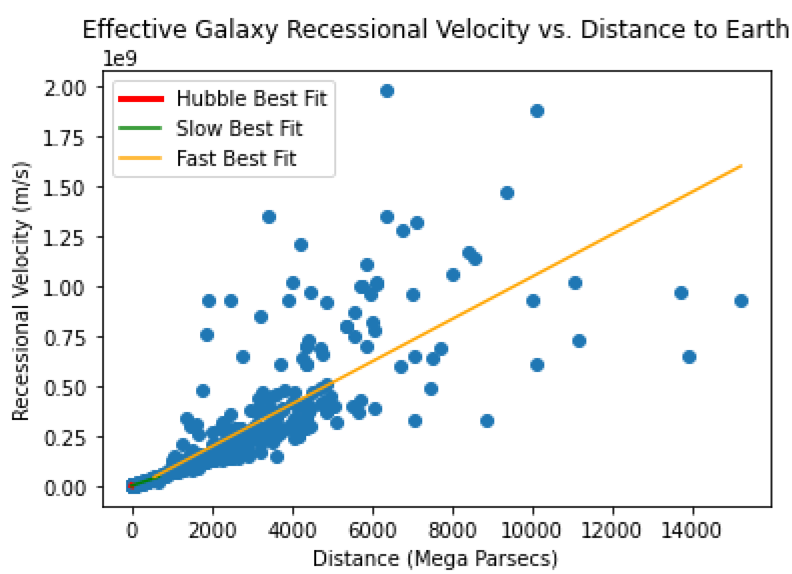


Therefore, according to this plot, recessional velocity appears to scale linearly with distance, but sort of drops off at larger distances.

Next, I calculated the effective recessional velocities for the three data sets by converting the tabulated velocity to a redshift using the relativistic Doppler formula. The meaning of these new velocities is the recessional velocities required to create the same Doppler shift if they were created only by a linear Doppler effect:



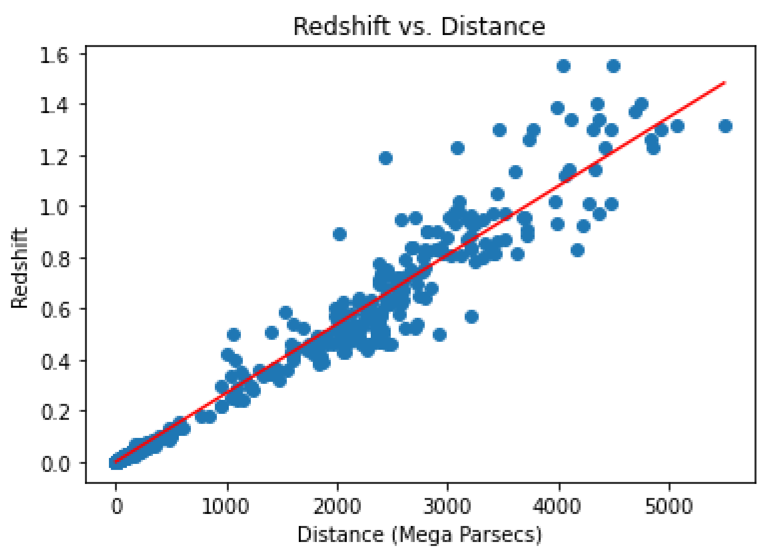
After calculating the new effective velocities, I overplotted the best-fit lines for the three data sets:



Recessional velocity is thought to scale linearly with distance at large distances, which is demonstrated in this plot because the best-fit lines slowly increase in slope.

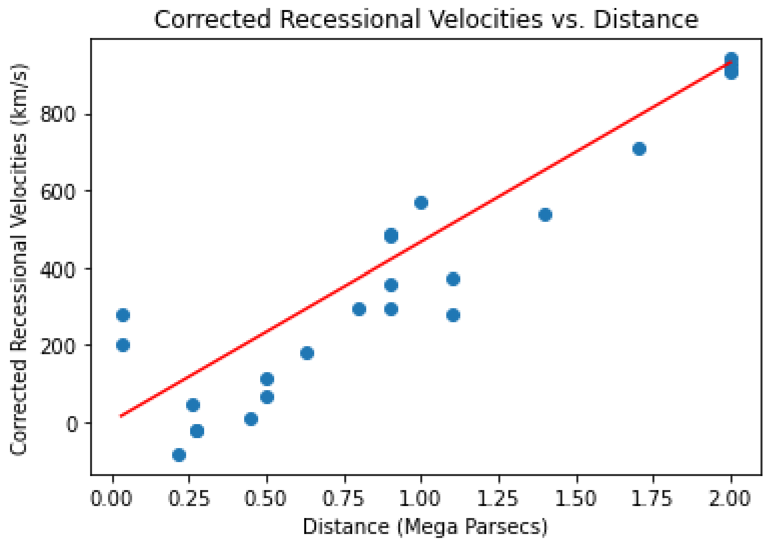
For the slow ($v < c / 8$) galaxy data, the Pearson correlation coefficient between the effective recessional velocity and the distance is $r = 0.97$ with $p = 0$. Therefore, with $r \sim 1$ and $p = 0$, this represents a very strong linear relationship between recessional velocity and distance. For the fast ($v > c / 8$) galaxy data, the Pearson correlation coefficient between the effective recessional velocity and the distance is $r = 0.79$ with $p = 3e\text{-}85$. Therefore, with $r = 0.79$ and $p \sim 0$, this represents a strong linear relationship between recessional velocity and distance.

Next, I made a scatter plot of Redshift vs. Distance for all the Type IA supernovae in the two modern data sets and overplotted the best-fit line:



This plot shows that galaxies further away from Earth are receding faster than galaxies closer to Earth, which is the main evidence for acceleration of the Universe.

Finally, I extracted the Right Ascension and Declination data from the Hubble csv file, and using the fit_correction function (from the class files) I corrected the velocities for solar motion. After scatter plotting the corrected velocities and overplotting the best-fit line, it looks less like the original plot that Hubble published (and my plot in part A):



While this plot looks much different from Hubble's, the Pearson correlation coefficient for this data is much better, at r = 0.93 and p = 9.4e-11. With r ~ 1 and p ~ 0, this indicates a strong linear relationship. My correction has improved my confidence in the linear portion of the fit, because the Pearson coefficient is significantly higher (compared to the original Hubble data in pat A.i) and the p-value is much lower, by a factor of about 1e-3. Furthermore, the variance and covariance for my corrected variables are also much lower than compared to the original Hubble data.

Thought Questions

1. Rather than taking one variable as dependent and the other as independent, another way to approach error analysis of two measured values is them both being dependent. Then, you could use the covariance of the measured values to analyze the error.
2. In part C, I found the Hubble's constant to be Ho = 465.9 km/s/Mpc and in part A, I found the Hubble's constant to be Ho = 454 +/- 75. Therefore, (465.9 - 454)/454 = 0.03 so the correction applied in part C changed my calculated Hubble's constant by about 3%, and the correction is well within the one-sigma uncertainty from part A.
3. The role of models and expectations in data analysis is to allow us to make predictions and test hypotheses. Say you have a predicted model for some natural phenomenon, you could use said model to test as much data as you want, which is where statistics tells us how well our predicted model fits the natural phenomenon. So Hubble's application of a correction was legitimate because he had good reason to try to correct for solar motion, which resulted in a more accurate Hubble constant with less uncertainty.
4. The age of the universe can be predicted using Hubble's constant, which has a current accepted value of ~70 km/s/Mpc. In this project, we calculated Hubble's constant to be about 470 km/s/Mpc, which is MUCH higher than the current accepted value. This discrepancy caused Hubble to calculate the age of the universe at around 2 billion years, which we now know is about 7 times too small. Hubble's contemporaries were worried about this implication of his measurements because it has large uncertainties, putting a large uncertainty on the age of the universe. Hubble's constant also has implications on the acceleration of the universe, so the large uncertainty puts a wide range on the amount of dark energy in the universe.

Comment

All good! Except there are quite a few grammatical/syntactic errors in the Project steps