Hubble constant Analysis

The aim of this report is to estimates of the expansion rate of the Universe, the Hubble constant.

Part 1 Cepheid period-luminosity (PL) relation

The aim of the of the first part of work base on data set 1(MW_Cepheids) is to find the α and β constant of the Milky Way Cepheids.

The α and β follows the Period-luminosity relation

$$M = \alpha * log_{10}P + \beta$$

$$(equation\ 1)$$

M = absolute magnitude(level of intrinsic brightness)
P = period of its pulsation
α, β = constant

The value of P is already obtained with low level of uncertainty which can be neglected, which just need to apply inside log function to get $log_{10}P$. The M needs to be found through the value of apparent magnitudes, m , and distance from galaxy to earth, d_{pc} and extinction, A. Where α, β are are required to get from the linear best fit from $M\ VS\ log_{10}P$

The relationship between M and m, d_{pc}, A named as Distance modulus

$$M = m - 5 * log_{10}d_{pc} + 5 - A$$

$(equation\ 2)$

M = absolute magnitude(level of intrinsic brightness)
m = apparent magnitudes(level of apparent brightness)
d_{pc} = distance from galaxy to earth(unit in pc)
A = extinction(intervening matter's dimming effect)

And d_{pc} is not provided in the data, and have to conclude from

$$d_{pc} = 1000/p_{mas}$$

$(equation\ 2.1)$

The Period-luminosity relation is a classic linear function, follows $y = m * x + c$, and the m and c value, in Period-luminosity relation will be α, β, are not independent to each other but rather correlated, in order to cancel this correlation, and find each one with a independent uncertainty, we have to modify this linear function,

$$y = m * x + c_1$$

as

$$y = m * (x - \bar{x}) + c_2$$

$\bar{x}$ = average of all value of x
c_2 = new intercept different than origin c_1

And will changed Period-luminosity relation to

$$M = \alpha * (log_{10}P - \overline{log_{10}P}) + \beta_2$$

$(equation\ 3)$

$\overline{log_{10}P}$ = average of all value of $log_{10}P$

β = new β(β_2) different than origin β

α = constant, remained unchanged

β_2 = constant, a new β that is not correlate to α

M finding

To calculated the value of M, absolute magnitude, The uncertainty of M need to be considered, and uncertainty of M is computed through uncertainty of other variables in $equation\ 2$, which are $m,\ d_{pc},\ A$.

Among these variables, m is provided with very accurate result, which its error can be neglected.

We calculated the uncertainty of M through two method, one is that we assume the the distribution of the uncertainty is following gauss distribution and done through numerical simulations of M of 100000 samples include $m,\ d_{pc},\ A$ with their uncertainty use gauss distribution for each object, and apply norm.fit to each of the gauss fit, norm.fit use maximum likelihood estimation, which the real value of M to be mean of the gauss distribution and error of the M to be standard deviation of gauss distribution.

Another method we have is using the error propagation formula to find the error of the M, and the Error propagation formula is

$$\sigma_f = \sqrt{(\frac{\delta f}{\delta a} * \sigma_a)^2 + (\frac{\delta f}{\delta b} * \sigma_b)^2 + (\frac{\delta f}{\delta c} * \sigma_c)^2 + ...}$$

We have conclude 2 list of error use 2 different method, which the results are

Object	M	M error(Gauss)	M error(Formula)
l-Car	-5.272019713	0.2279556726	0.2242429666
zeta-Gem	-3.92877602	0.1446581264	0.1437639035
beta-Dor	-4.01435176	0.1221702339	0.1214209245
W-Sgr	-3.913325765	0.1949532321	0.1928280169
X-Sgr	-3.638393726	0.1638662525	0.1642408376
Y-Sgr	-3.285101983	0.3075938188	0.2983401417
delta-Cep	-3.452594573	0.0937671087	0.09391522371
FF-Aql	-3.0244684	0.1522605056	0.1514866968
T-Vul	-3.194231995	0.2743784827	0.2696231952
RT-Aur	-2.834943791	0.1909798368	0.1896112876

And we can clearly see that the a error concluded by the gauss distribution use numerical simulations and the error computed by the error propagation formula showed a high level of similarity, which we can reassured to use numerical simulations base on the assumption that error distribution is following gauss distribution for any further work about uncertainty calculation.

α and β finding

While the values of M has been calculated, $curve\_fit$ from $scipy$ routine has been used to fit the data to the idea model of Period-luminosity relation($equation\ 3$). Which used a non-linear least squares to find the best parameters of a function.

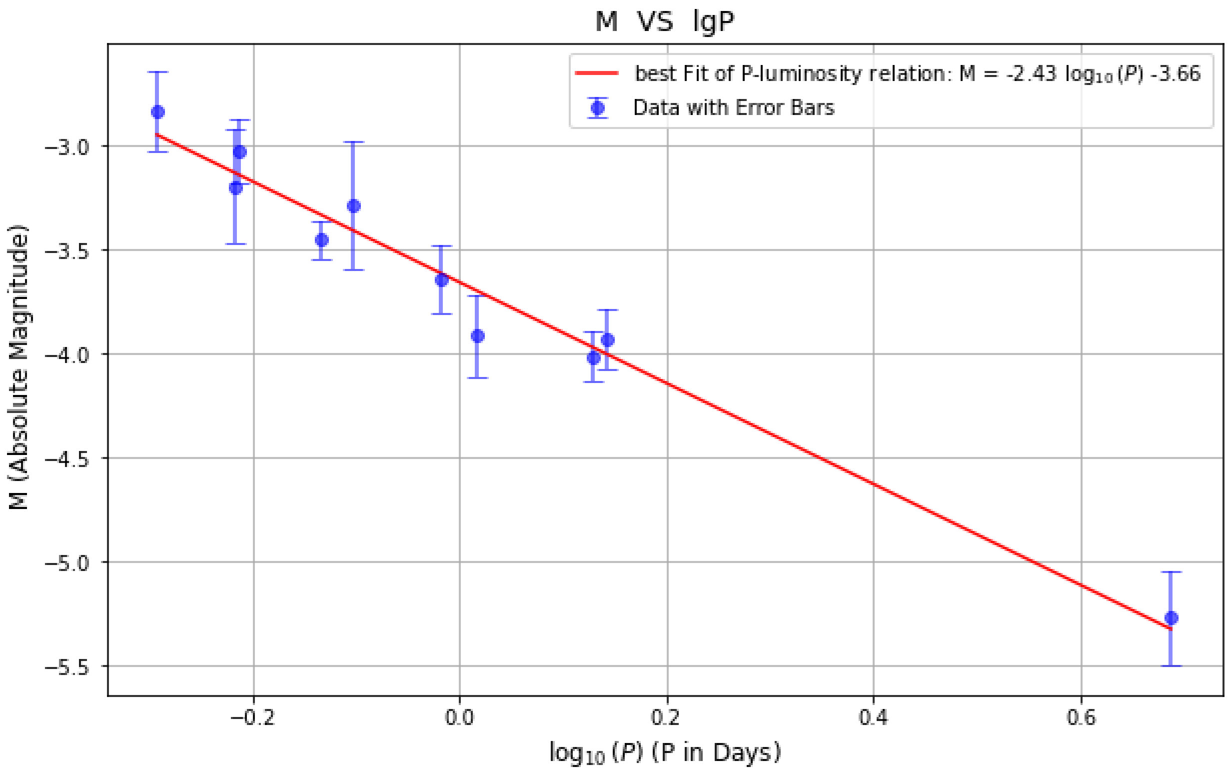


Figure 1: Absolute magnitude(M) VS the Log P. The observed values are plotted with the blue points with error bars. The estimated value is plotted by the red line.

With the curve fit function, it will provide with both the value of error of , x value of the function with the fitted data, in Period-luminosity relation will be α and β. The covariance matrix, will provide both the uncertainty of α and β, and the correlation level between α and β.

$$\alpha = -2.43 \pm 0.143$$
$$\beta = -3.66 \pm 0.038$$

correlation level between α and $\beta \approx 0.15\,\%$

The relationship between the α and β is showed to be lower than 0.5 %, which is small enough to be neglected in the further calculation, and can be treat as independent for each data.

Which is indeed a low value of correlation, and can be neglected for further calculation involve α and β.

The above Figure 1 showed the line of best fit of the Period-luminosity relation and the provided data in data set 1(MW_Cepheids).

We can see from Figure 1 that all data points with error bars have included the estimated value(best fit line) within their uncertainty range, which shows that this is a greatly fitted model.

And provided with a chi-square value,

$$X_v^2 = 4.507,$$

and for the model that have 8 degree of freedom, it has a critical value of 15.507, which $critical\ value\ >\ X_v^2$, and is much higher than the X_v^2 the value of the current set of data fitted with the Period-luminosity relation, and we can conclude that the data is fitted well with the Period-luminosity relation model, and we can use this model for the further calculation.

But reduced X_v^2 has provide a value around 0.56 which implies that the error of the current measured data might have been overestimated in the dataset 1(MW_Cepheids) .

Part 2 distance of a nearby galaxy

The aim for the 2nd part of the work is to find distance from galaxy ngc4527 to earth with Mpc unit, $d(Mpc)$, based on data set 2(ngc4527_cepheids.dat) and previous calculated value of α, β. Also $\overline{log_{10}P}$ from 1st part is a constant offset for new $log_{10}P$ of this data set 2.

In order to calculate the $d(Mpc)$, we need rearrange the $equation\ 1$, $equation\ 2$, and convert form pc to Mpc, we got

$$d(Mpc) \ = \ 10^{(\alpha*(log_{10}P - \overline{log_{10}P_{old}}) + \beta - m + A - 5\ )/\ (-5)} / \ 10^{6}$$

$(equation\ 4)$

$\overline{log_{10}P_{old}}$ = mean of $log_{10}P$ from previous equation 3, base on data set1(MW_Cepheids)

We then computed the result of distance, $d(Mpc)$, of each cepheid from earth based on the assumption that error distribution is following gauss distribution, and used numerical simulations with gauss distribution of 10000 samples follows $equation\ 4$, include all variable α , β, m, $log_{10}P,\ A$ with their given uncertainty, and applied norm.fit to each computed gauss distribution of cepheid, which the value of $d(Mpc)$ and uncertainty of the $d(Mpc)$ to be mean and standard deviation of each corresponding gauss distribution respectively.

And use $np.average$, a function that can calculate mean with weights of uncertainty and conclude a weighted mean of data, to all the $d(Mpc)$ we have got, then we have find the value of $d(Mpc)$ to be

$$d(Mpc) \ = \ 14.82 \pm 0.65\ Mpc$$

And because these Cepheids are all from the same galaxy(*ngc4527*), which ideally they all should have a same distance, $d(Mpc)$, from the earth, and by plot $d(Mpc)$ VS $log_{10}P$ graph

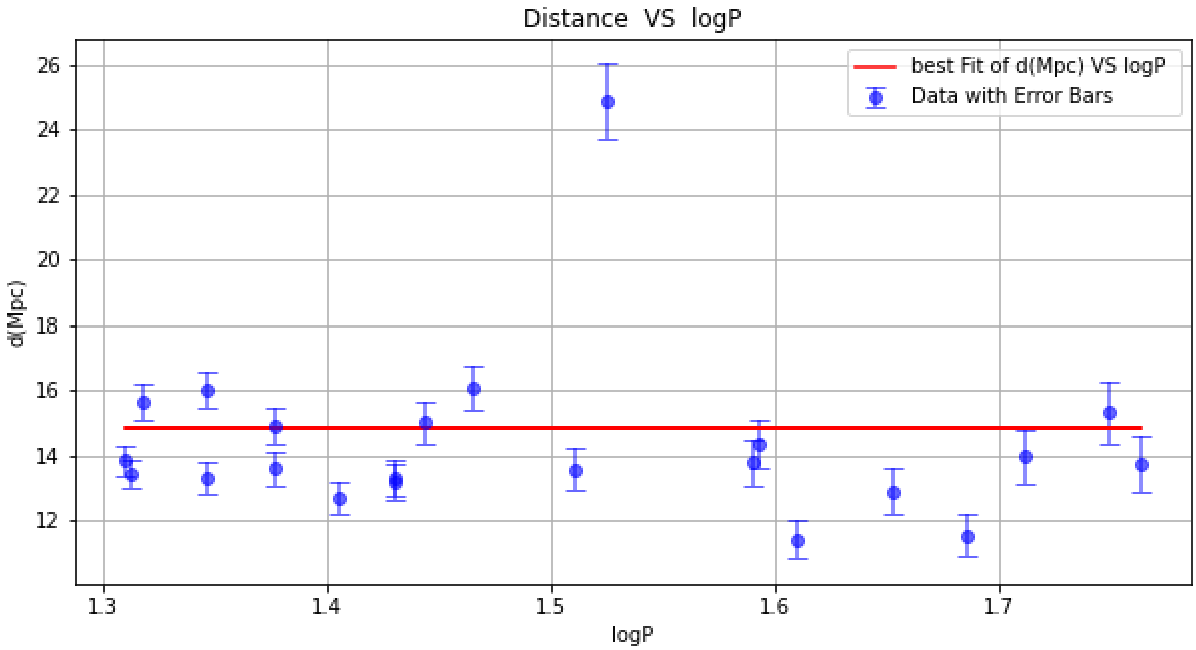


Figure 2: Distance(Distance from galaxy to earth) VS Log P(P = period, in days). The observed values are plotted with the blue points with error bars. The estimated value is plotted by the red line.

We can see in Figure 2, data point with error bar is mostly not covered predicted value within their uncertainty range, means that the value of estimated $d(Mpc)$ not fit well with the actual observed data.

And the value of chi-square we have got for *Figure 2* is,

$$X_v^2 \approx 223.099$$

with 21 degree of freedom, critical value to be 32.671 which

$$X_v^2 > critical\ value,$$

And this shows that the data is not very correlated to the ideal estimated value of data, and with

$$\text{reduced } X_v^2 \approx 10.596$$

Which is much bigger than the 1, and means that either the estimated value does not fit well to the observed value or the actual uncertainty is underestimated.

But we can see that there is a obvious outlier in *Figure 2* that stay much away from the estimated value of $d(Mpc)$(fit line), and if we move it away, chi-square value and reduced chi-square reduced to

$$X_v^2 \approx 148.802$$

$$\text{reduced } X_v^2 \approx 7.086$$

Which still

$$X_v^2 > critical\ value$$

and reduced X_v^2 from larger than the 1, but in a significantly improved scale, and is more acceptable.

Better value

We have been noticed that, ideally, Cepheids in NGC4527 may be affected from extra extinction, A, but estimated error has not been provided for this data set.

In order to offset this extra intrinsic error, median is always shows better stability to stay near to real mean of data, and median instead of average has been used for each variable against the shift of the data, M, A due to the extra intrinsic error contained in each variable.

Follows the $equation\ 4$, each data of $\alpha, \beta, m, A, log_{10}P$ all have certain level of uncertainty on the data, and can be displayed in the form $data \pm \sigma(data)$, but the uncertainty of $m, log_{10}P$ is slight enough to be ignored, and error of extinction, A, in the data has offset for a significant proportion by using the median of all variables.

We assumed the error is distributed following gauss distribution again, and a numerical simulation of gauss distribution of 10000 samples follows $equation\ 4$ with same procedure has gone through for these median values in order to find a better mean and error of d(Mpc)

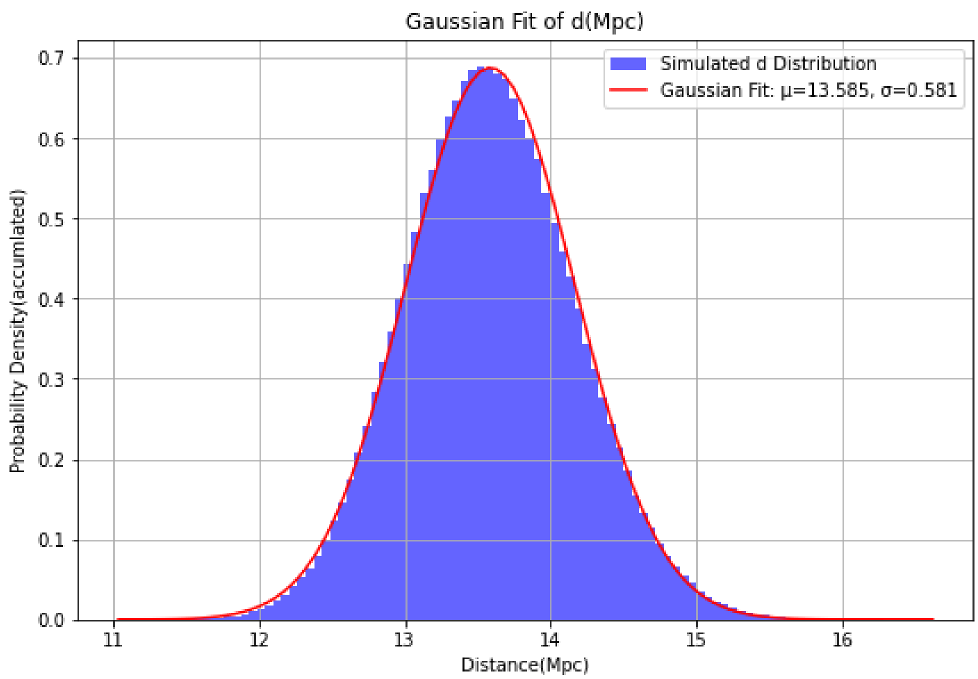


Figure 3: Absolute magnitude(M) VS the Log P. The simulated values are plotted with the blue column. The estimated value(fit line) is plotted by the red line.

Then the value of d(Mpc) we have got from this norm fit is

$$Median\ d(Mpc):\ 13.57 \pm 0.58\ Mpc$$

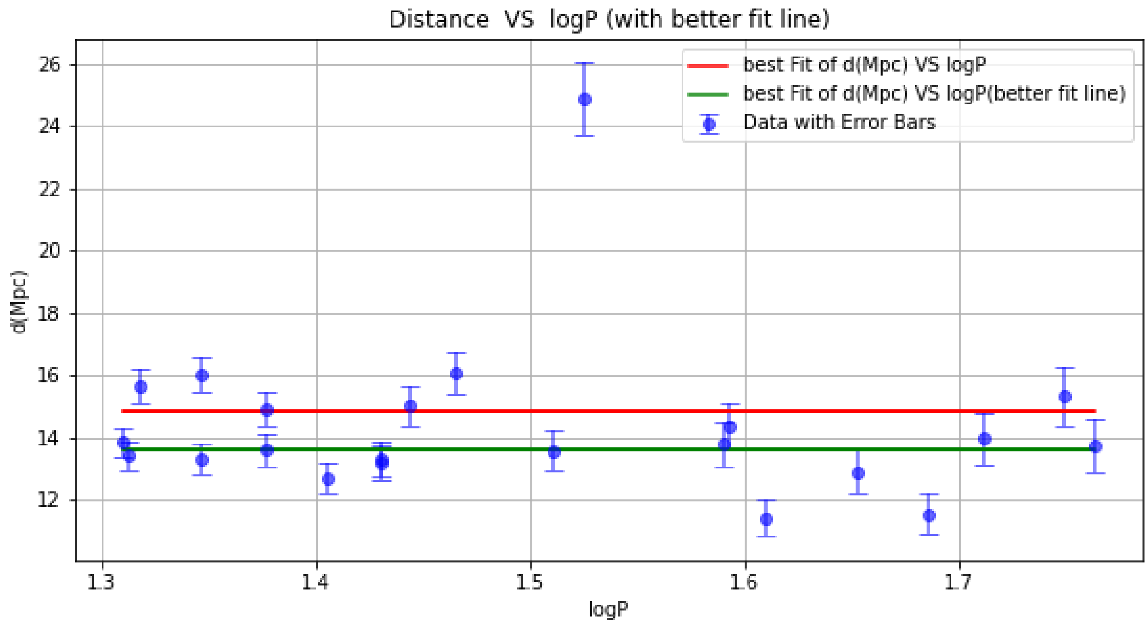


Figure 2.1: Distance(Distance from galaxy to earth) VS Log P(P = period, in days). The observed values are plotted with the blue points with error bars. The origin estimated value is plotted by the red line. The new estimated value is plotted by the greenline.

And the new chi-square and reduced chi-square for d(Mpc) VS $log_{10}P$ are

$$X_v^2 \approx 108.398$$

$$\text{reduced } X_v^2 \approx 5.16$$

Which has significant decreased, better stilled indeed showed that the error of the data has extremely underestimated, and hardly to further higher the fit level of model. With the offset to the extra extinction of the d(Mpc), we are use this new result of d(Mpc)

$$Median\ d(Mpc) = 13.57 \pm 0.58\ Mpc$$

as the optimal value to use in further calculation.

Part 3 Hubble Constant

The aim for the 3rd part of the work is to find hubble constant of the universe, based on data set 3(*other_galaxies.dat*) and 1 previous calculated value of d(Mpc) of *ngc4527* with a provided corresponding recession velocity of $1152\ km/s$.

The hubble constant follows hubble law,

$$v = H_0 * D$$

(equation 5)

v = recession velocity
H_0= Hubble constant
D = distance from galaxy to earth

And can be rearrange to calculate Hubble constant, H_0

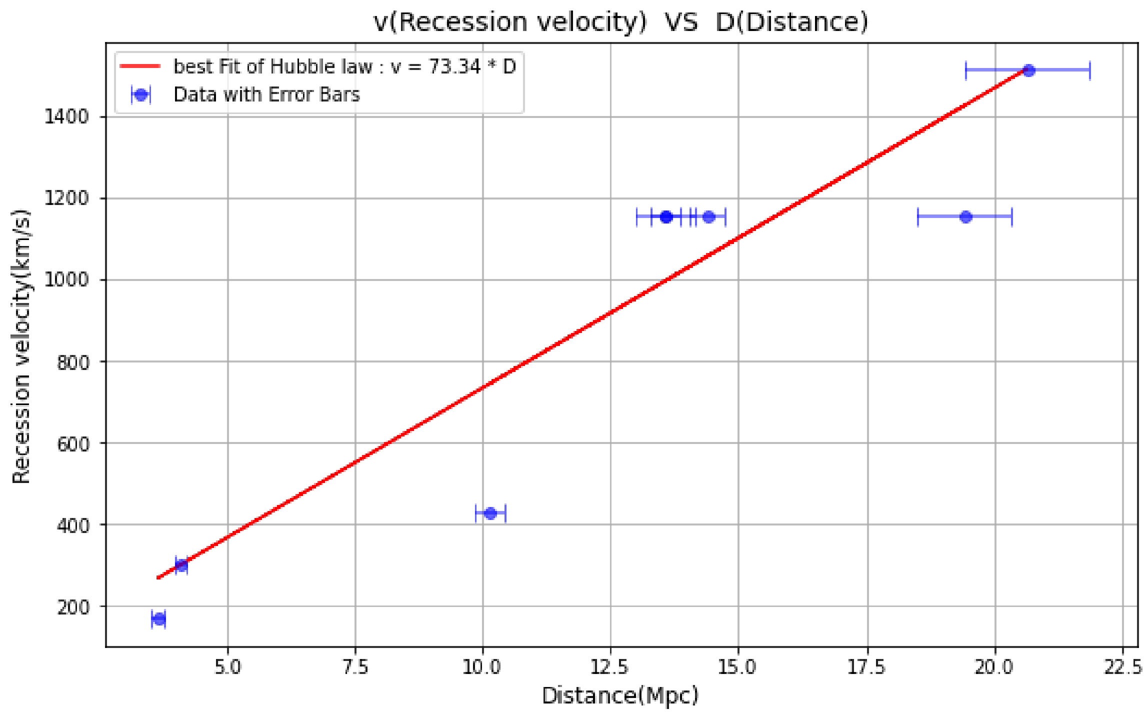
$$H_0 = \frac{v}{D}$$

(equation 6)

The values of D, v, error of D are provided, where recession velocity, v, are provided with high levels of accuracy with neglectable errors.

The value of the Hubble constant has been calculated using $equation\ 6$, and for better display of error, and the value of error of D has added to each data points in graph of D VS v, instead of v VS D, since that only D have significant error. We then used curve_fit from scipy routine to fit the data to the ideal model of *equation 6.*

Origin graph (v/D)(not good to see error)



New graph (D/v)(good to find error)

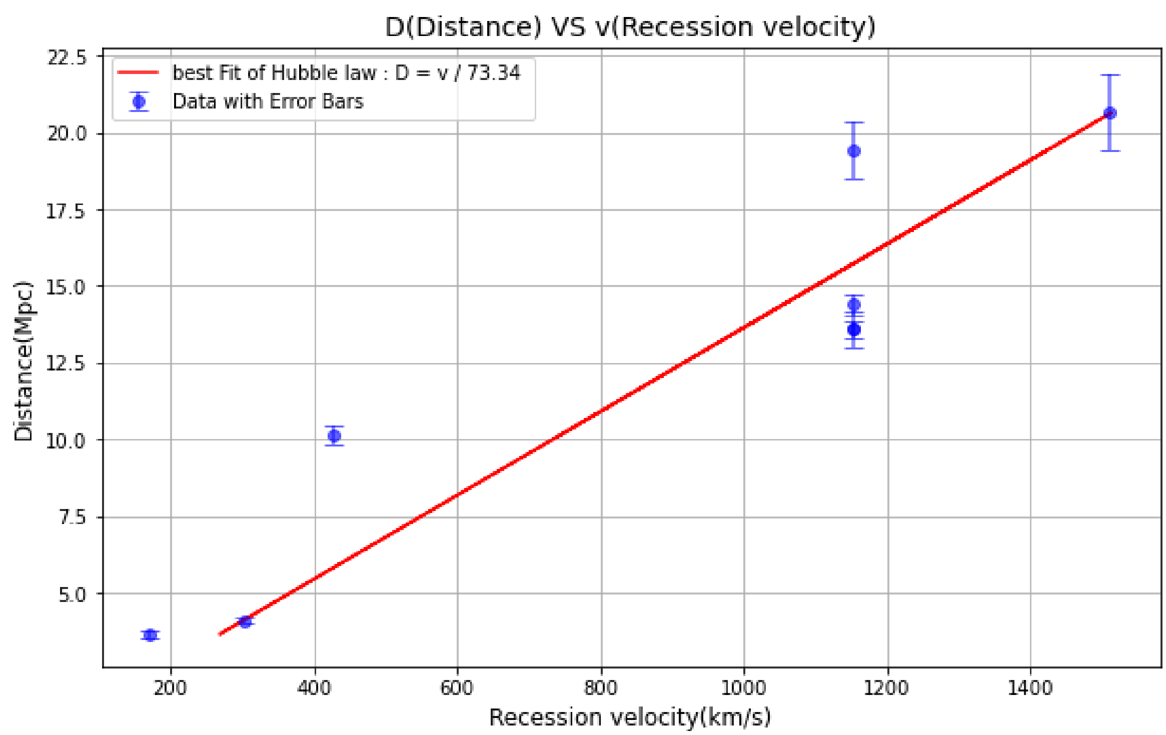


Figure 4: Recession velocity VS Distance(Distance from galaxy to earth). The observed values are plotted with the blue points with error bars. The estimated value is plotted by the red line.

Figure *4* has a significant number of points laid out of the expected value(best fit line) of the data, and mostly not covered the expected value within their uncertainty range. Which means that the value of estimated Hubble constant does not fit well with the actual observed data.

And the hubble constant we have got

$$H_0 = 73.34 \pm 23.53\ km\ sec^{-1} Mpc^{-1}$$

The value of chi-square and reduced chi-square we have got for *Figure 4* is,

$$X_v^2 \approx 426.585$$

reduced $X_v^2 \approx 60.941$

with 7 degree of freedom, critical value to be 14.067 which

$$X_v^2 > critical\ value$$

reduced $X_v^2 >> 1$

Which means the estimated data is highly uncorrelated with the actual observed value, and the error of data is highly underestimated.

There is another possibility the ideal model is not very correct, and we have see that the data has shown an interesting sin-like trend, which showed a constant up and down trend, but with different periods for each cycle, this implies there is a reduced coefficient in this trend of behaviour.

Some of this behaviour might be due to the extra extinction, but apparently most of this still hints that a new model is required for a better prediction, but more data is required to get the ideal model of this new trend.

Better value

In order to find a more averaged and precise value to cancel extra intrinsic error, median is required to find the real mean of data of a skewed data.

A numerical simulation of gauss distribution of 10000 samples follows $equation\ 6$ has conducted for each galaxy.

And then sum all the results of these numerical simulation value, and divide by the number of data we have simulated, and due to the data with low error is more concentrated than the data with high error, this new averaged hubble constant will shift more to the real value of true hubble constant if the error providing is correct.

And based on the current result, our optimized result for hubble constant is

$$H_0 = 68.16 \pm 2.41\ km\ sec^{-1} Mpc^{-1}$$