STA130 Capstone Project

Poster

Presentation

Riyad | Nicolas | Henry | Chloe

3 Research Questions Includes Data, Method, Visulization & Conclusion

STA130 Capstone Project

Poster

Presentation

Riyad | Nicolas | Henry | Chloe

3 Research Questions Includes Data,

Method, Visulization & Conclusion

RESEARCH QUESTION 1 - How is the galaxy's total size related to the percentage of light within the radius?

**Author - Nicolas Dias Martins**

INTRODUCTION

There are three different variables related to a galaxy's size in the Galaxy Zoo Tabular Data: petro_th50, petro_theta, and petro_th90. Our goal with this research question is to find understand how this variables are related.

DATA

'Galaxy Zoo Tabular Data Contents' (galaxies' sizes in radius): petroth50 (50% of the light inside the radius), petro_theta (total size), and petro_th90 (90% of the light inside the radius).

METHODOLOGY

Bootstrapping to get 3 different simulate sampling distributions, one for each variable beforementioned and compare to see if there is any relation between them.
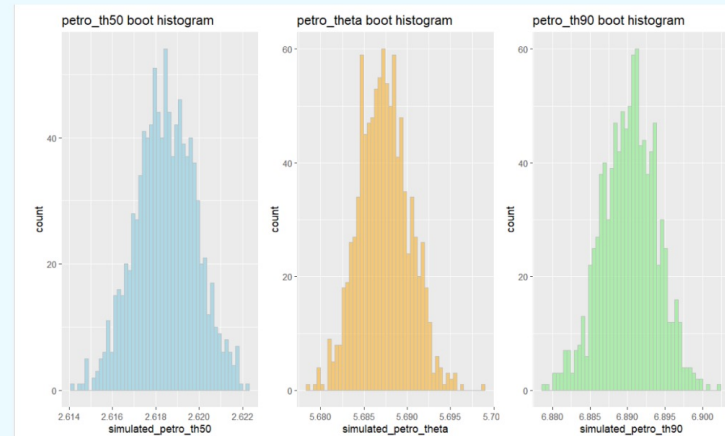
CONCLUSION

There is a clear relation between a galaxy's total size and the percentage of light within the radius (between 50% and 90% of light within the radius).

'DO THE GALAXIES FARTHEST TO THE EARTH DIFFER SIGNIFICANTLY FROM THE CLOSEST ONES IN TERMS OF THEIR TOTAL LUMINOSITY?'

answered, using hypothesis test

by Riyad Valiyev
TUT0101 Group #6

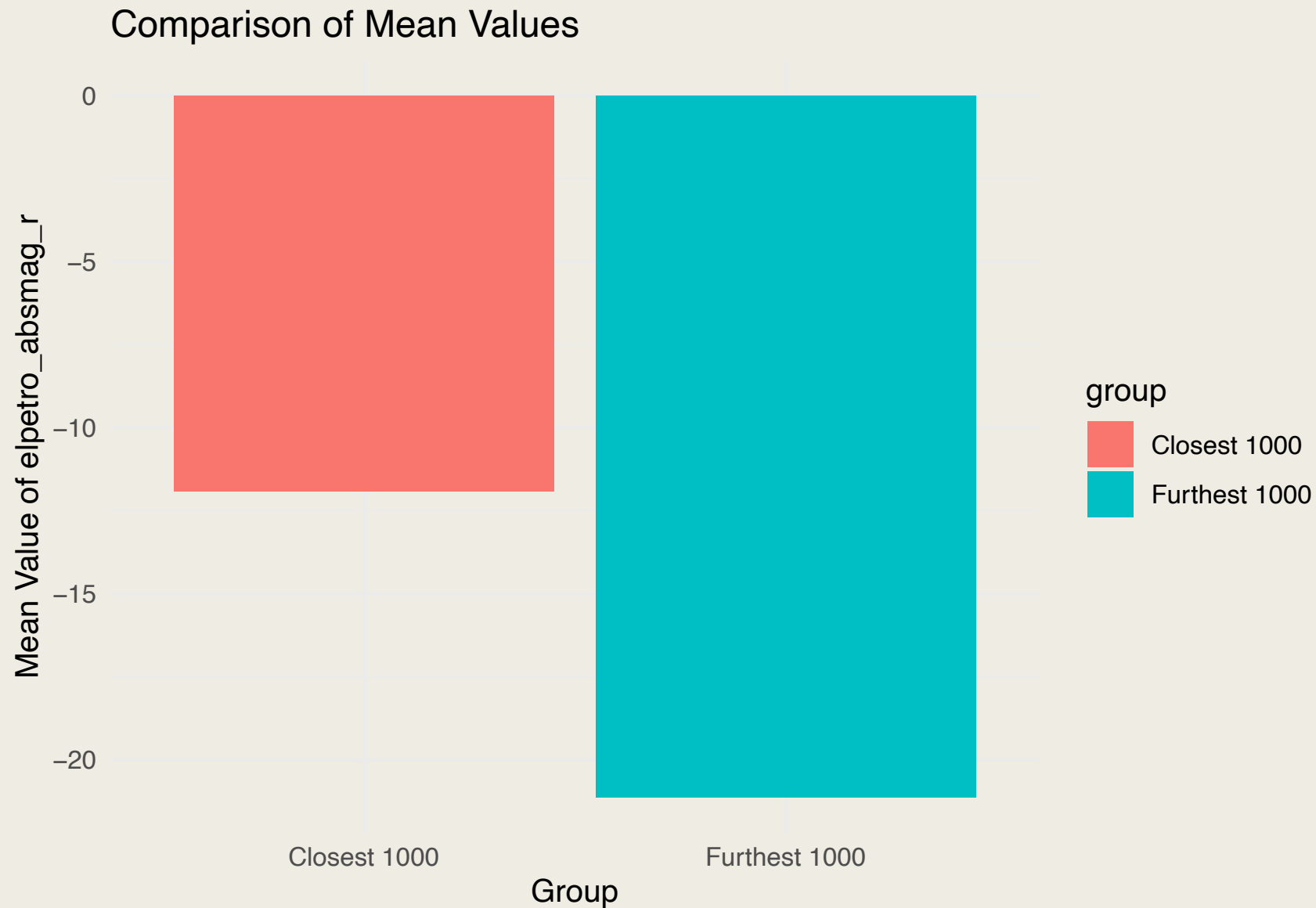# Samples of the furthest and closest galaxies

```
## Rows: 1,000
## Columns: 3
## $ elpetro_absmag_r <dbl> -21.93583, -21.48143, -19.69226, -21.43395, -22.41922~
## $ redshift         <dbl> 0.1499996, 0.1499991, 0.1499987, 0.1499983, 0.1499975~
## $ distance         <chr> "furthest", "furthest", "furthest", "furthest", "furt~
```

The ***furthest*** 1000 galaxies with their associated redshift and total brightness values

```
## Rows: 1,000
## Columns: 3
## $ elpetro_absmag_r <dbl> -0.8244038, -1.1099548, -6.2743492, -12.0312986, -9.2~
## $ redshift         <dbl> -0.0044815368, -0.0035639217, -0.0033348468, -0.00312~
## $ distance         <chr> "closest", "closest", "closest", "closest", "closest"~
```

The ***closest*** 1000 galaxies with their associated redshift and total brightness values

Correlation between redshift and total brightness

o Arranged 641,166 galaxies in an ascending order of their redshift values.

o Clustered them in the groups of 1000.

o Computed the mean "total brightness" of each group.

o Removed the outliers (only kept the groups with means in IQR).

o Plotted the result with a line graph.
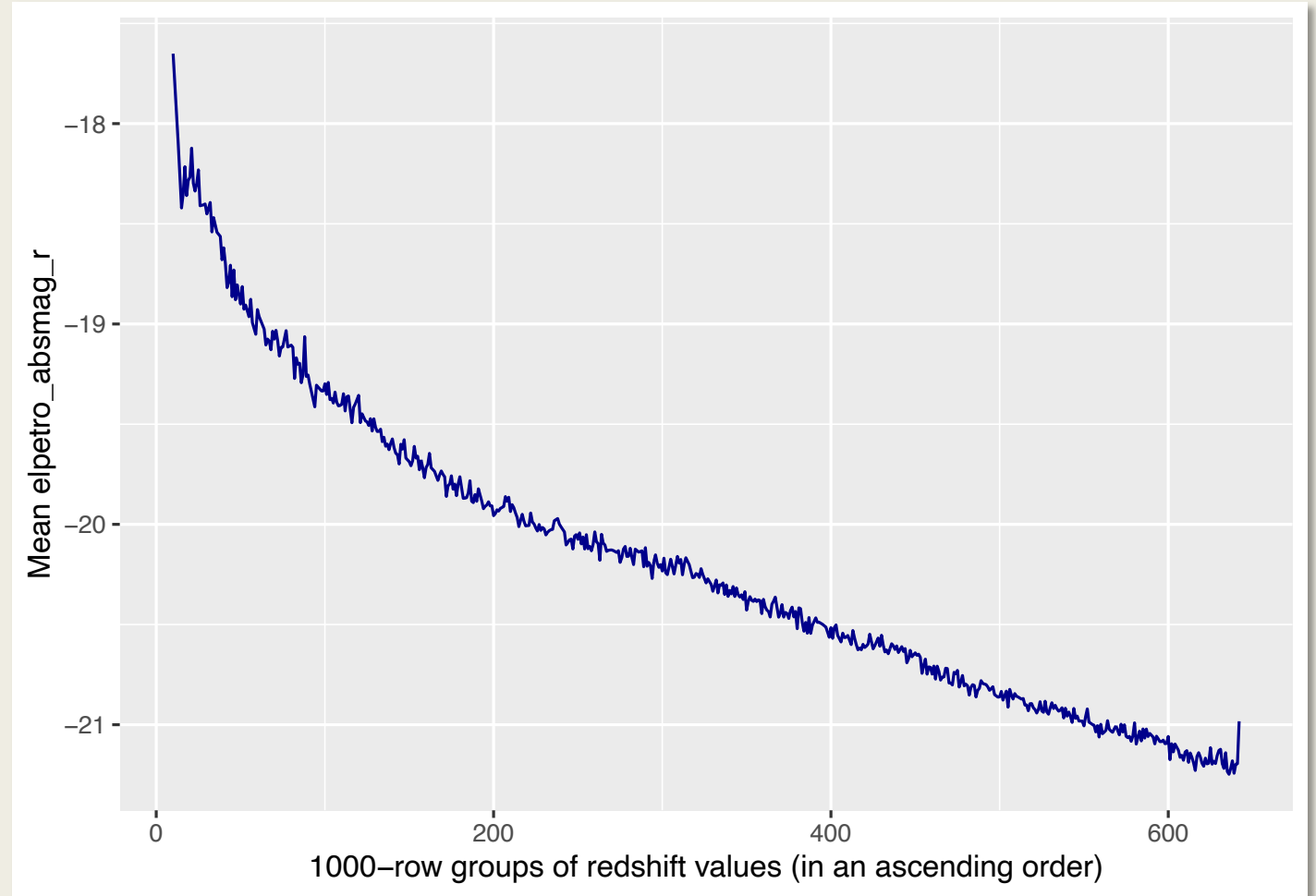
Galaxy groups with higher redshift have less total brightness.

Hypothesis Test Results: Zero p-value!
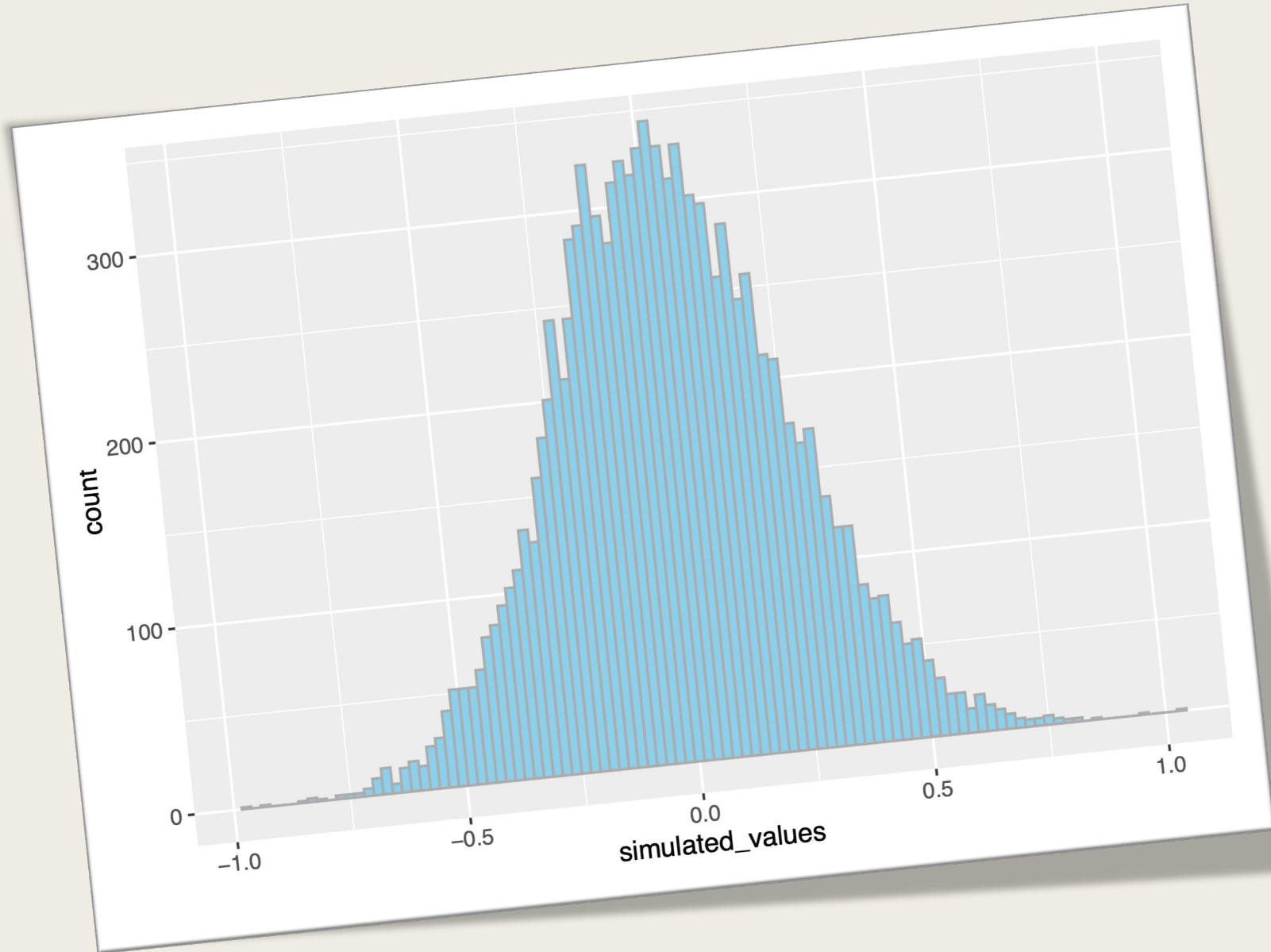
RESEARCH QUESTION 3

Do the galaxy's apparent brightness and the redshift linearly correlated?

Authors
Henry & Chloe

The research is the capstone project poster presentation in the course STA130

INTRODUCTION

In GalaxyZoo Tabular Data content, there are two variables that are related which are redshift and elpetro_absmag_r.
Therefore, we decide to investigate do the galaxy's apparent brightness and redshift linearly correlated?

VISUALIZATION

N

`Fitted line plot & Histogram

METHODOLOGY

The method we used in research question 3 is simple linear regression. Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous or quantitative variables.
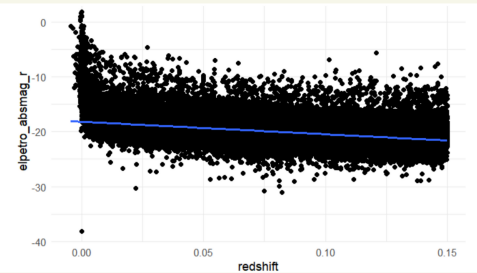
DATA

`Galaxy Zoo Tabular Data Contents'
redshift - related to how far away that galaxy is from us;
elpetro_absmag_r - An estimate of the galaxy's total luminosity brightness or intrinsic brightness measured in absolute magnitude

ANALYSIS

Firstly, we update the data by loading libraries ('tidyverse' and 'arrow') and remove all the 'NA' values.
Secondly, we create a linear regression plot , which 'redshift' is the x-axis and 'elpetro_absmag_r' is the y-axis.
Thirdly, we create a 'mod' the analysis the relationship between 'redshif' and 'elpetro absmag r'. We apply 'summary$coefficients' to find the B1 , B0 and p-value.
Fourly, we can do hypothesis test . H0 will be there is a linear relationship between brightness and redshift . H1 will be there is not a linear relationship between brightness and redshift . The p-value we got form summary will tell us whether to reject H0 .

CONCLUSION

The B0 is the brightness when redshift equal to zero, which is -18. B1 is the average change in brightness for 1 unit change in redshift which is - 13.3
Form here, we can get the linear regression model:

$$y_i = \beta_0 + \beta_1 x_i$$
$$y_i = -18 - 13.3 x_i$$

$\alpha = 0.05$. The p-value is 0.11 which is bigger than α , we will reject H1, support H0.
There is a linear relationship between brightness and redshift.

RESEARCH QUESTION 3

How well can a unary linear regression model predict the galaxy's apparent brightness from redshift?

Authors

Henry & Chloe

The research is the capstone project poster presentation in the course STA130

INTRODUCTION

In GalaxyZoo Tabular Data content, there are two variables

VISUALIZATION

`Fitted line plot & Histogram

DATA

`Galaxy Zoo Tabular Data Contents'

In GalaxyZoo Tabular Data content, there are two variables that are related which are redshift and elpetro_absmag_r. Therefore, we decide to investigate how well can a unary linear regression model predict the galaxy's apparent brightness from redshift?
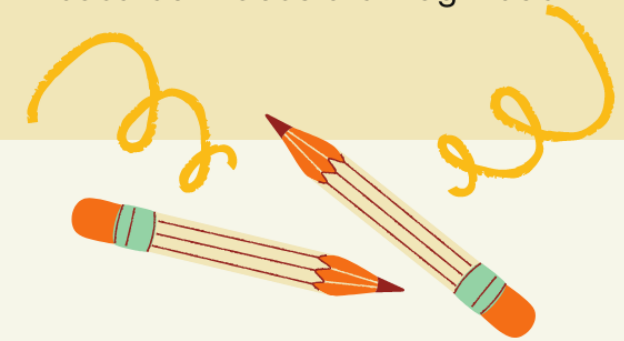
`Fitted line plot & Histogram

The method we used in research question 3 is simple linear regression. Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous or quantitative variables.
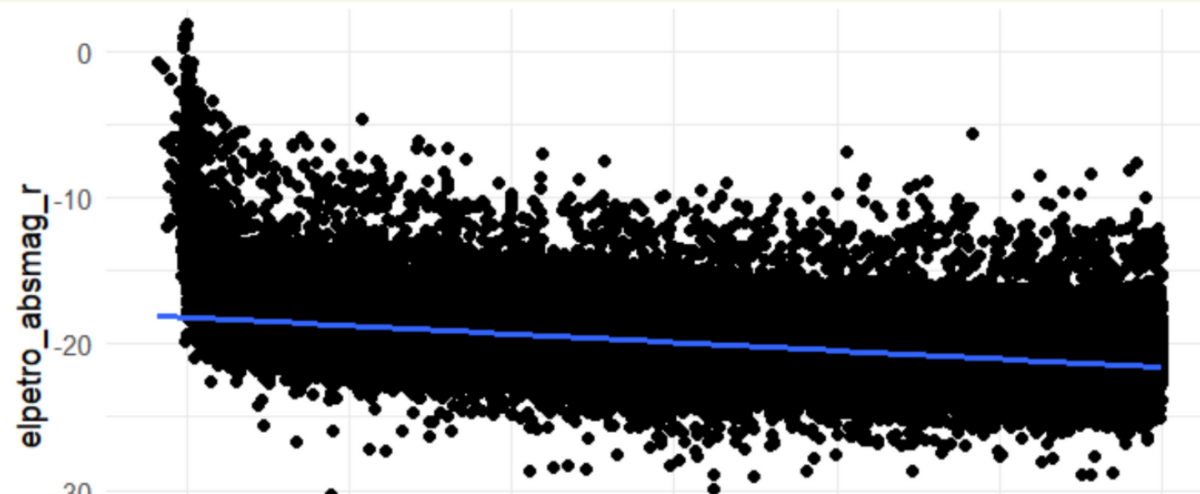
`Galaxy Zoo Tabular Data Contents'
redshift - related to how far away that galaxy is from us;
elpetro_absmag_r - An estimate of the galaxy's total luminosity brightness or intrinsic brightness measured in absolute magnitude
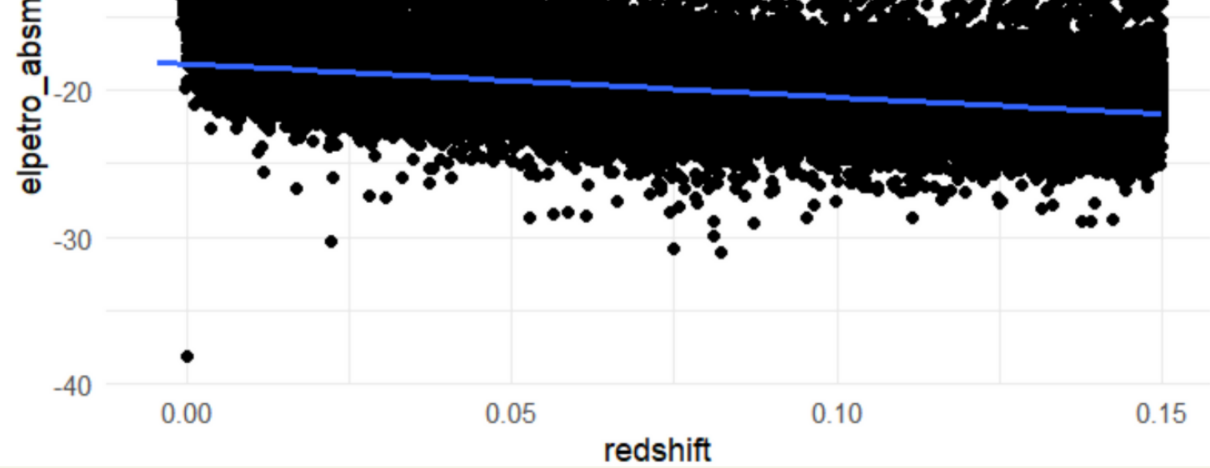
Firstly, we update the data by loading libraries ('tidyverse' and 'arrow') and remove all the 'NA' values.
Secondly, we create a linear regression plot , which 'redshift' is the x-axis and 'elpetro_absmag_r' is the y-axis.
Thirdly, we create a 'mod' the analysis the relationship between 'redshif' and 'elpetro absmag r'. We apply 'summary$coefficients' to find the B1 , B0 and p-value.
Fourly, we can do hypothesis test . H0 will be

which 'redshift' is the x-axis and 'elpetro_absmag_r' is the y-axis.

Thirdly, we create a 'mod' the analysis the relationship between 'redshif' and 'elpetro absmag r'. We apply 'summary$coefficients' to find the B1 , B0 and p-value.

Fourly, we can do hypothesis test . H0 will be There is a linear relationship between brightness and redshift . H1 will be There is not a linear relationship between brightness and redshift . The p-value we got form summary will tell us whether to reject H0 .

CONCLUSION

The B0 is the brightness when redshift equal to zero, which is -18. B1 is the average change in brightness for 1 unit change in redshift which is - 13.3

Form here, we can get the linear regression model:

$$y_i = \beta_0 + \beta_1 x_i$$
$$y_i = -18 - 13.3 x_i$$

$\alpha = 0.05$. The p-value is 0.11 which is bigger than α , we will reject H1, support H0.

There is a linear relationship between brightness and redshift.