ST511 Project Report

Vijay Tadimeti

[tadimetv@oregonstate.edu](mailto:tadimetv@oregonstate.edu)

**Introduction:**

For this project I have decided to use the IMDB Top 250 Movies database. It contains columns/variables that include Rank, Rating, Title of the movie and the number of Votes. The data includes a list of 250 movies each of which has a rating out of 10.

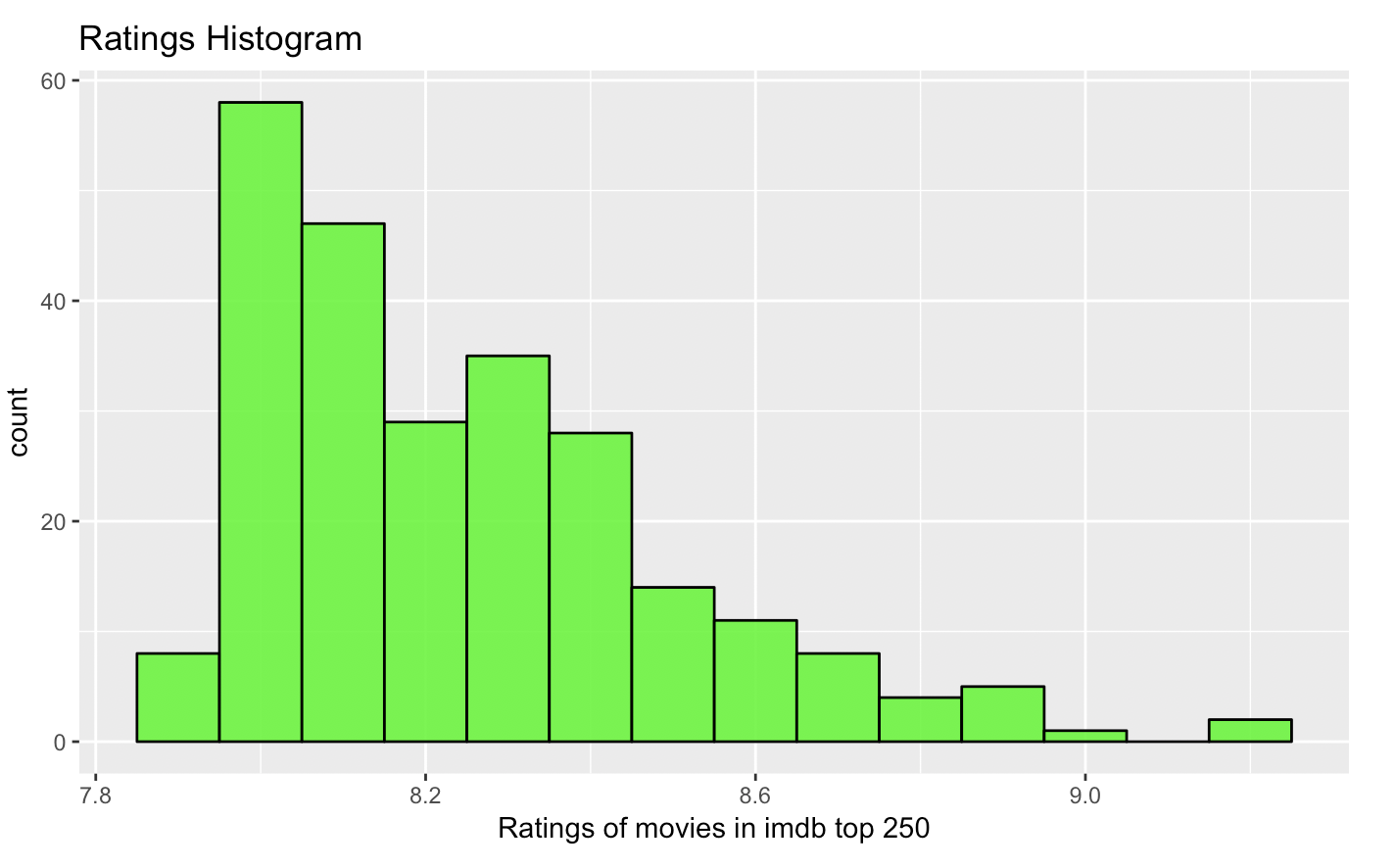
Population: 250 films that meet the eligibility criteria

Variable of Interest: In this dataset we'll look at the IMDB User Score instead. Possible values are between 0 and 10.

Scientific question: Is the average IMDB rating equal to 8.2 for movies in the top 250 list?

Statistical translation: “Is μ = 8.2 ?”

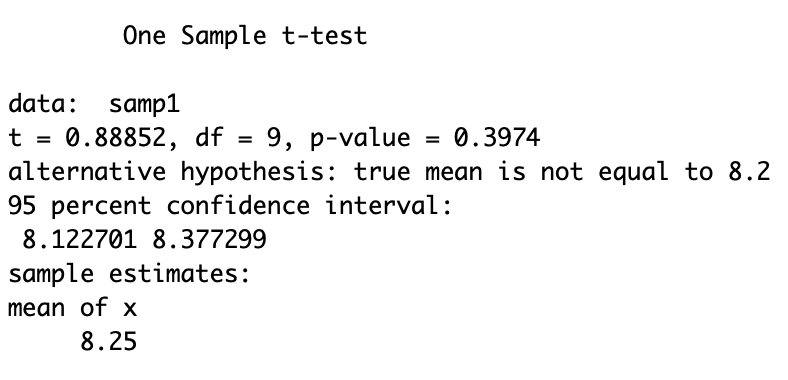
Exploratory Plot of the data: Here, I have chosen to use a histogram to describe the data.



**Methods:**

I have chosen to perform a T test on this data. I am assuming that I do not know the population variance σ2. To be precise I’m using a One sample t-test. I have observed the following T statistic, and p value, both of which are mentioned below in the results section.

**Results:**



I have observed that the T test has provided me with the following values.

T statistic = 0.88852, p = 0.3974 at a confidence interval of 0.95.

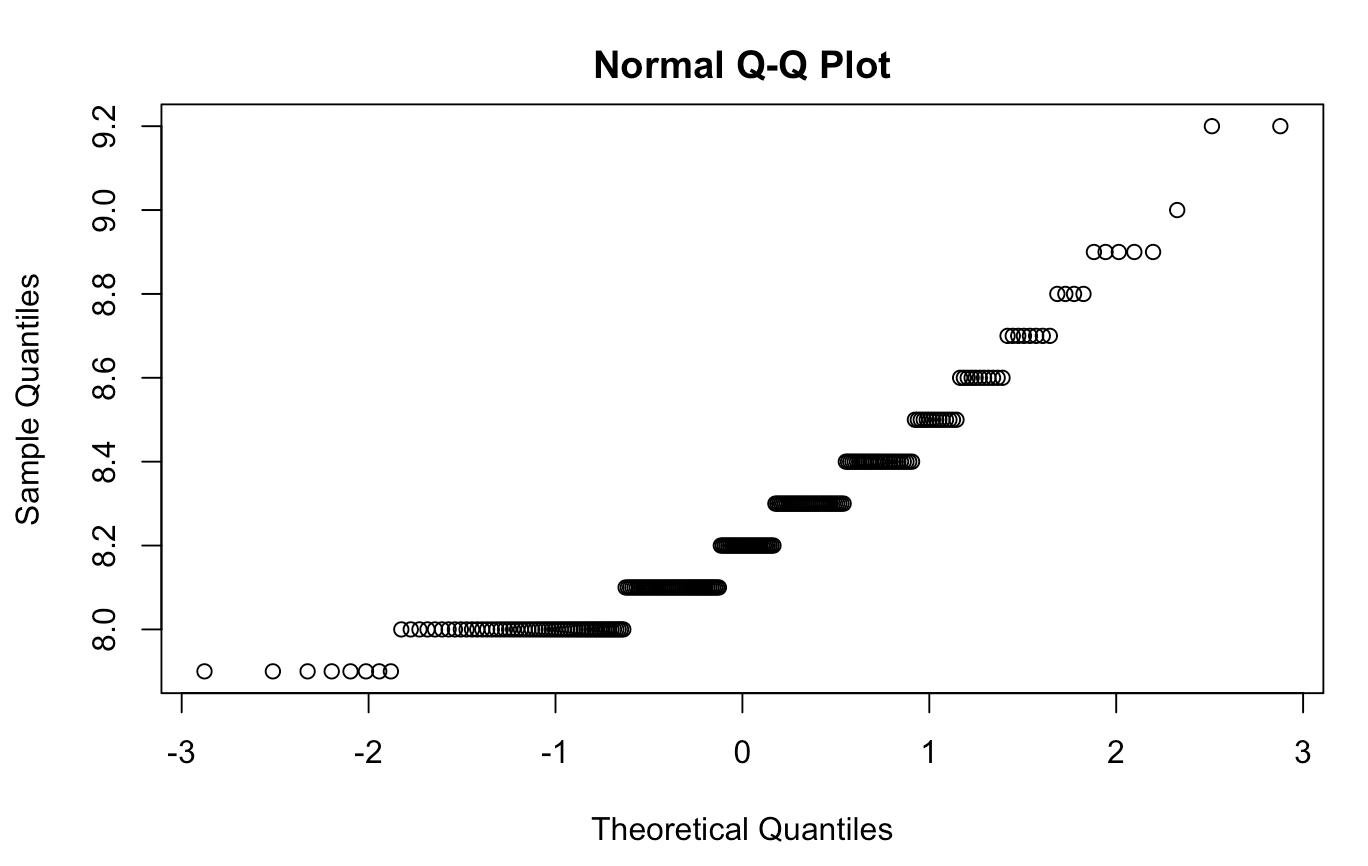
**Assessment:**

1. The statistical independence of the observations within each sample.

2. The Normality of the underlying population distributions.

To assess assumption 1 we have to think carefully about how the data were obtained. Typically, data that are collected close together in time and/or space, or data that are collected together in some kind of cluster. In this case the data was not obtained in this fashion.

To assess the Normality of the underlying population distributions, we have to take a look at the shapes of the sample distributions and make a judgement call about whether you think the samples could plausibly have come from Normal populations. Here I have chosen a Q-Q plot. The general outline of the plot below suggests that it is of linear nature.

****

**Conclusion**:

We fail to reject the null hypothesis $Ho:\mu=8.2 in favor of alternate hypothesis at significance level alpha=0.05, since our p-value is greater than 0.05.