1. **Problem Statement:**

The random dataset of 100 Fiction and Non-fiction books are given to check if the readers like Fiction books or Non-fiction books based on the average of ratings. Mostly any avid reader has its own preference of book taste. So, here checking that by the ratings can give us some insights for Book market.

1. **Data Description:**

The dataset used for the analysis consist the list of Fiction and Non-fiction Books procured from the internet. Here, the data has been cleaned and separated in two different categories namely, Fiction and Non-Fiction and any book that had non number ratings or prices been excluded. Also, out of number of books only 100 books are taken into consideration based on random selection.

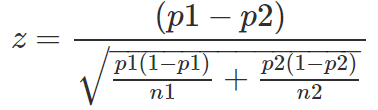
1. **Hypothesis Testing Procedure:**

Here, the data consist rating from out of 5. Based on that we will be using Z-test and to get better understating of the result we will be performing Two-sample Proportional Test. For the analysis the ratings are given for each book out of 5. Firstly, I need to divide ratings of the books with 5 and then find the average of the proportional data. In the data, we are looking for the best ratings of the group that is why proportional test will be more useful for the analysis.

* Here, the Null hypothesis **H0** will be, Asses weather there is a significance difference in the proportional ratings of the Fiction and Non-fiction books.
* Alternative hypothesis **Ha** will be, that there is no significance difference in the proportional ratings of the two groups.

Alpha level will be used as 0.05.

Two-Sample Proportional Test:

The equation for the analysis is:

Here, p1 is the average proportional of Fiction books p1 = 0.9148

p2 is the average proportional of Non-Fiction books p2 = 0.925

n1 is number of samples for Fiction books n1 = 100

n2 is number of samples for Non-Fiction books n2 = 100

The value of Z after calculation will be, z stats = -0.2657.

The critical value of Z from Z table based on α is, z critical = ±1.96.

P value upon calculation is p = 0.79047.

1. **Hypothesis Testing Result:**

Here z statistical is -0.2657 & z critical is ±1.96. Now, the statistical z lies in the area of curve of critical z. Hence, we fail to reject the Null hypothesis. The amount of reader that like Fiction more than Non-fiction is cannot be said based on the analysis.

Also, the Pearson correlation coefficient of the books prices is 0.103.

1. **Summary and Conclusion:**

From the analysis we can finally say that there cannot be seen a difference between readers who like Fiction more than Non-fiction. The test shows that type or genre of the books doesn’t matter when it comes to reading. Also, the correlation between both groups is positive but almost 0. So, we can say that the price for the both groups does not vary much so both types of books are easily available to reader.

1. **“What I learned” Statement:**

Firstly, I would like to say that understanding statistic is much harder for the real time problems than it is in textbooks problems.

The amount of time to understand the data and to which analysis to use and if it will be able to give me the result I am looking for or not is really frustrating. But still doing all the hassle is always worth it in the end. Because even if I don’t get the desire output that I am looking for, I still learn something new or at least which test to use based on the data I have been given, after a lot of practice.

The thing that I understood the most is before tyring to do any analysis; I should look for all the values that I have available and do I have enough data to perform the type of analysis based on those data and if it will give me better output or not.