ANALYSIS OF ANIMATED MOVIES

CHILDRENS' VS. NON-CHILDRENS'

Steve Newns

RESEARCH QUESTION

Is there a significant difference between the ratings of those animated movies created for children and those animated movies that were *not* created for children?

MOTIVATION

It seems to be a popular preconception that animated forms of media are deemed to be mainly (or only) for children. This includes such things as comic books, television cartoons, video games, and animated movies.

I was interested in looking at different animated movies and analyzing how animated movies targeted to both children and non-children alike were rated, and seeing if there was anything interesting in the data that would suggest that animated movies targeted towards non-children were of a high quality, which could be used as an argument against the previously-mentionted preconceptions of animated media.

DATASET

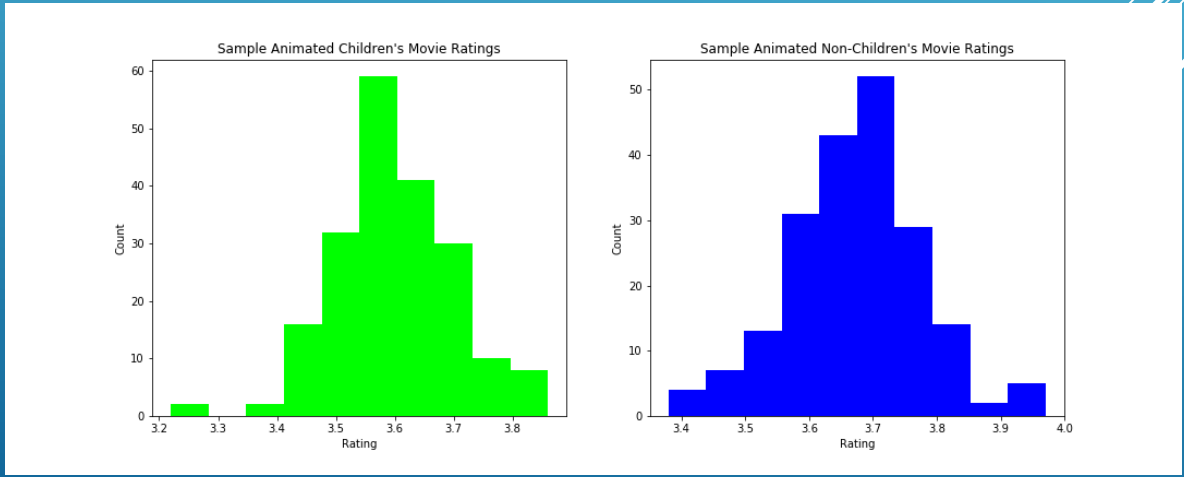
- The dataset used for this analysis IMDB Movie Dataset
- This dataset was created from the MovieLens web site (http://movielens.org), a recommendation site, via GroupLens Research
- The data were collected between 1995 and 2015 and contains 20 million ratings between 1 and 5 stars, as well as 465,000 tags applied to 27,000 movies by 138,000 users.

FINDINGS

The original dataset was subsetted into only animated movies, and then this animate movies dataset was itself subsetted into 2 new subsets, those with "Children" as one of the genres they were categorized as, and those who were not categorized as so.

The datasets were not normally distributed, so a sampling distribution of 200 means from each dataset was created to use for the analysis in a t-test, which assumes normal distributions are being used.

As we can see, both sampling distributions were relatively normal, and the non-children's animated movies sampling distribution seemed to have a higher average than the children's' animated movies distribution.



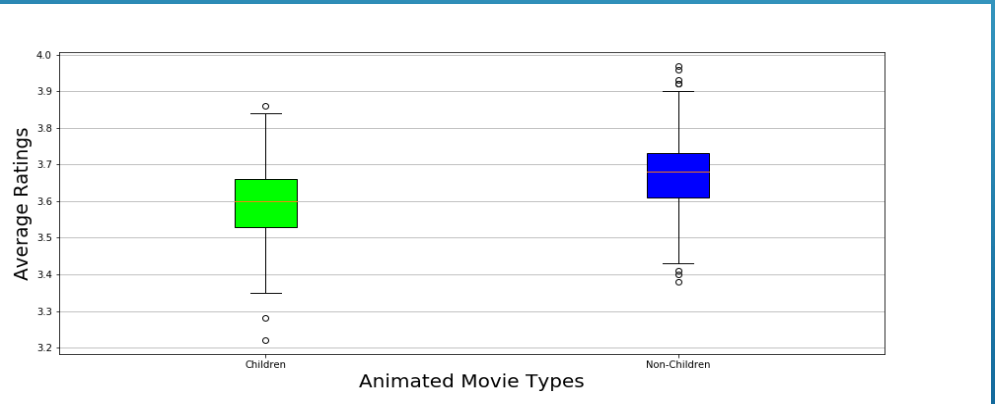
FINDINGS

Looking at boxplots was another way to see that non-childrens animated movies seemed to be more highly rated, but I had to make sure this difference was significant

I chose to do a one-tailed, independent samples t-test comparing both samples I had taken with the sample statistic displayed below, to the right of the boxplots

My null hypothesis [**H(0)**] was that the difference in mean animated movie ratings was 0 or greater than 0 (i.e. ratings were not significantly different OR the mean animated children's movie rating was significantly higher than the mean animated non-children's movie rating)

My alternative hypothesis[**H(A)**] was that the mean non-children's animated movies rating was significantly higher than the mean children's animated movies rating




```
Mean of Sample of Animated Children's Movies: 3.6
Standard Deviation of Sample of Animated Children's Movies: 0.102

Mean of Sample of Animated Adult Movies: 3.67
Standard Deviation of Sample of Adult Children's Movies: 0.104

Degrees of Freedom: 398
```

FINDINGS

The critical value here for a confidence level of 95% ($\alpha = 0.05$) for this one-tailed independent samples t-test* with degrees of freedom above 120 was -1.645 (check with a t-table)

The t-test value I ended up with when subtracting the animated non-children's movies sample mean from the animated children's movies sample mean and dividing by the standard error was -9.59, which fell far into the critical region (was much less than the t-critical value), so I was able to reject null hypothesis.

With this value, I could state with 95% confidence that animated non-children's movies were rated significant higher than animated children's movies

With this information, one could go back to other forms of animated media and do some similar analysis to bolster the idea that animated media is not just for children.

* one-tailed because I was just looking for a difference between the sample means in a certain direction