Modeling and prediction for movies

Amin Ghaderi

February 15, 2018

Setup

Load packages

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
library(statsr)
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.4.3
```

Load data

We use load command to import the movies data.

```
load("movies.Rdata")
```

Part 1: Data

Acquisition: This data is randomly selected from IMDB and Rotten Tompato APIs from movies produced before 2016.

Population: To be included in this data set, the movie needs to be (1) in the Rotten Tomatoes and IMDB databases, (2) produced before 2016.

Causality/Generalization: Since the data is randomly sampled from the discussed population and no ***random assignment*** is performed, the results of this study does not demonstrate any causality. Any results could be merely used to demonstrate correlation. The results is also only generalizable to the poplation discussed above, which are movies in IMDB and RT databases, produced before 2016. * * *

Part 2: Research question

In this study, we are choosing our target (dependant variable) as the IMDB score (`imdb_rating`). We would like to see how this score is affected by different factors. We are mostly interested in the effect of `critics_rating`, `audience_rating`, and whether movie is on Top 200 Office Box list (`top200_box`). We will also study the effect of rest of the parameters and finally will build a model that can predict the IMDB populaity of the movies based on the available parameters.

Part 3: Exploratory data analysis

NOTE: Insert code chunks as needed by clicking on the "Insert a new code chunk" button above. Make sure that your code is visible in the project you submit. Delete this note when before you submit your work.

Part 4: Modeling

NOTE: Insert code chunks as needed by clicking on the "Insert a new code chunk" button above. Make sure that your code is visible in the project you submit. Delete this note when before you submit your work.

Part 5: Prediction

NOTE: Insert code chunks as needed by clicking on the "Insert a new code chunk" button above. Make sure that your code is visible in the project you submit. Delete this note when before you submit your work.

Part 6: Conclusion