Project 1 Instructions

Niklas Cypris

6. 4. 2021

Bechdel Data Set

https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-03-09

Bechdel Test

Introduced by graphic novelist Alison Bechdel in a 1985 strip from her "Dykes To Watch Out For" series.

Figure 1: Original Comic Strip, https://en.wikipedia.org/w/index.php?curid=41021832

3 Criteria

1. At least two women in the movie

2. They have a conversation with each other
3. The conversation isn't about a male character

Think of your favorite movies - how many of them pass that test?

Research Questions

For movies that were released from 1990:

RQ1: Does a movie passing the Bechdel test predict its overall box office success?

RQ2: Does critical acclaim predict a movie's overall box office success?

RQ3: Does a movie's budget predict its overall box office success?

RQ4: Does a movie's success depend on any of the above-mentioned factors working in concert?

Tasks

- Read data set into workspace (*session 3*)
- Only keep movies that were released from 1990 onwards (*session 4*)
- Visualize missing cases (*session 3*)
- Compute a fail_pass variable from the clean_test variable (*session 4*)
  - Compute a varibale that is called fail_pass and has the values 0 and 1 from the variable clean_test that has the values *dubious, men, notalk, nowomen,* and *ok*. The last value means that a movie passed the Bechdel test.
- Compute gross overall income of movie - select 2013 levels (*session 4*)
  - Compute a variable that is called *gross* per movie which is the sum of the intgross_2013 and domgross_2013 variables.
- Mean center the predictors *pass_fail, budget_2013, metascore,* and *imdb_rating* (*session 7*)
- Visualize distributions of the metric predictors *budget_2013, metascore,* and *imdb_rating* (*session 5*)
  - Create histograms for each of the variables that display their distributions
- Visualize distributions of pass-fail by year (*session 5*)
  - Create barplots for the absolute and for the relative frequency of movies that passed and that failed the Bechdel test by year.
- Compute a multiple regression with *gross* as the criterion and *pass_fail, budget_2013, metascore,* and *imdb_rating* as well as their interactions as the predictors. (*session 7*)
- Check if the assumptions for a multiple regression are fulfilled (*session 7*)
  - If the VIF is too high, consider computing a variable that is called *acclaim* which is the mean score of the z-standardized *metascore* and *imdb_rating* variables

  - If any of the plotting methods that were introduced in *session 7* do not work, try the `plot()` function

- Decide if a parametric method or a non-parametric bootstrapping procedure should be conducted to answer the research question and apply it (*session 7*)
- Write an answer to the research questions