Lab 5

Group Member Names - here

Lab Overview

All students attending class in the group can turn in a single document with each participants name. Students not attending class will need to complete their own lab.

Download the movies earnings dataset. In today's lab, we are interested in using the budget of a film to estimate the earnings.

```
movies <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/movies_earnings.csv')
```

```
## Parsed with column specification:
## cols(
##   title = col_character(),
##   budget = col_integer(),
##   revenue = col_integer(),
##   release_date = col_date(format = "")
## )
```

```
movies
```

```
## # A tibble: 5,380 x 4
##    title                                    budget   revenue release_date
##    <chr>                                     <int>     <int> <date>
##  1 Star Wars: The Force Awakens            245000000    2.07e9 2015-12-15
##  2 Titanic                                 200000000    1.85e9 1997-11-18
##  3 The Avengers                            220000000    1.52e9 2012-04-25
##  4 Jurassic World                          150000000    1.51e9 2015-06-09
##  5 Furious 7                               190000000    1.51e9 2015-04-01
##  6 Avengers: Age of Ultron                 280000000    1.41e9 2015-04-22
##  7 Harry Potter and the Deathly Hallows: ~ 125000000    1.34e9 2011-07-07
##  8 Frozen                                  150000000    1.27e9 2013-11-27
##  9 Beauty and the Beast                    160000000    1.26e9 2017-03-16
## 10 The Fate of the Furious                 250000000    1.24e9 2017-04-12
## # ... with 5,370 more rows
```

1. (5 points)

Before looking at the data, summarize the relationship you expect between the budget of the film and the earnings.

2. (5 points)

A SRS of 100 movies has been taken for you from the dataset. Create a plot of earnings vs. budget for these movies. How does this figure match your expectations?

```
movies_sample <- movies %>% sample_n(100)
```

3. (5 points)

Report a point estimate of the population ratio B. Where does this show up on the graphic you created for problem 1?

4. (5 points)

Assume that you know the population mean for the movie budget,

```
xbar_U <- movies %>% summarize(mean(budget)) %>% pull()
```

using the 100 samples, calculate the mean earnings per film with both the SRS estimator and the ratio estimator.

5. (5 points)

Discuss which esimator from part 4 you prefer.