Lab 5: Top Grossing Films

Testing (One Group Means)

- Week 3: Hypothesis Testing (Two Group Means)
- Week 4: Hypothesis Testing (Categorical Data)
- **Week 5: Hypothesis Testing (More Than Two Group Means)**

Readings
Reading Check due May 03, 2016 at 17:00 UTC

Lecture Videos
Comprehension Check due May 03, 2016 at 17:00 UTC

R Tutorial Videos

Pre-Lab
Pre-Lab due May 03, 2016 at 17:00 UTC

Lab

Like most Americans, people in Austin are fascinated with cinema. The American film industry has captured the attention of audiences around the world, making film a multibillion-dollar-a-year industry. Most of the top-grossing films of all times have been produced by the same five major studios: 20th-Century Fox, Paramount, Sony Pictures, Universal Pictures and Warner Bros. This data set focuses on the 151 films made by these studios that made the list of the 245 top-grossing films of all times, as determined by source Box Office Mojo. For each of the films, data includes film genre, MPAA rating, measures of film critic and user rankings, and production outcomes such as budget, time in theaters and amount grossed.

(2/2 points)

Review of ANOVA

In this lab, you will use **ANOVA** to answer a question of interest. Let's start by remembering why we use ANOVA.

1a. What is the goal of an **ANOVA** analysis?

- (•) to determine if significant mean differences exist between multiple groups ✔
- ( ) to identify the distribution of counts across three or more groups
- ( ) to compare the variability of scores in two different groups

Lab due May 03, 2016 at 17:00 UTC

Problem Set
Problem Set due May 03, 2016 at 17:00 UTC

1b. Two specific group means can be said to be **significantly different** if:

- the F statistic of the overall ANOVA is significant
- one group mean is at least twice the value of another
- a Tukey HSD pairwise comparison shows $p < 0.05$ (or the identified level of significance) ✔

You have used 1 of 1 submissions

(1/1 point)

Lab Preparation

In this lab you will be working with data from the top grossing films of all time.

1. Open RStudio. Make sure you've installed the SDSFoundations package.
2. Type **`library(SDSFoundations)`**. This will automatically load the data for the labs.
3.Type **`film <- FilmData`**. This will assign the data to your Workspace.

Alternatively, you can use the steps in the "Importing a Data Frame" R tutorial video, and use

the FilmData.csv file. (Right-click and "Save As.") Make sure to **name** the dataframe "film" when importing.

1. Open RStudio.
2. Click on the "Import Dataset" button at the top of the workspace window. Choose *"from text file."*
3. Click on the location of the FilmData.csv file you just downloaded.
4. Click on the FilmData.csv file. Then, click Upload.

Feel free to use the script from the week's PreLab, which you can modify for use in this Lab.

2. **Two** of the following questions will be answered in this lab using **ANOVA**. Select the questions that can be answered with this method.

- [x] Which studio(s) earn a greater percentage of their earnings domestically? ✔
- [x] Which studio(s) are more successful in keeping their films in the theaters longer? ✔
- [ ] Which is a better predictor of how much a film will gross: the film's budget or how long it was in theaters?

✔

You have used 1 of 1 submissions