

UTAustinX: UT.7.20x Foundations of Data Analysis - Part 2



- Important Pre-Course Survey
- Contact Us
- How To Navigate the Course
- Discussion Board
- Office Hours
- Week 0: Introduction to Data (Optional Review)
- Week 1: Sampling
- ▶ Week 2: Hypothesis

Week 5: Hypothesis Testing (More Than Two Group Means) > Pre-Lab > Examine the Data

■ Bookmark

Reflect on the Question

Analyze the Data

Draw Conclusions





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.

Primary Research Questions

- 1. Does a film's rating (PG, PG-13, or R) impact its cost to produce?
- 2. Does a film's rating (PG, PG-13, or R) influence its IMDB score?

(1/3 points)

Check the Data

Let's begin by examining our data in R.

- 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.
- 4. Look at the spreadsheet view of the data to answer the following questions.

Lab due May 03, 2016 at 17:00 UTC

Problem Set
Problem Set due May 03, 2016

at 17:00 UTC

Alternatively, you can follow 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.
- 5. Look at the spreadsheet view of the data to answer the following questions.
- 1a. What numerical rank does Titanic hold among highest-grossing films?



Answer: 2

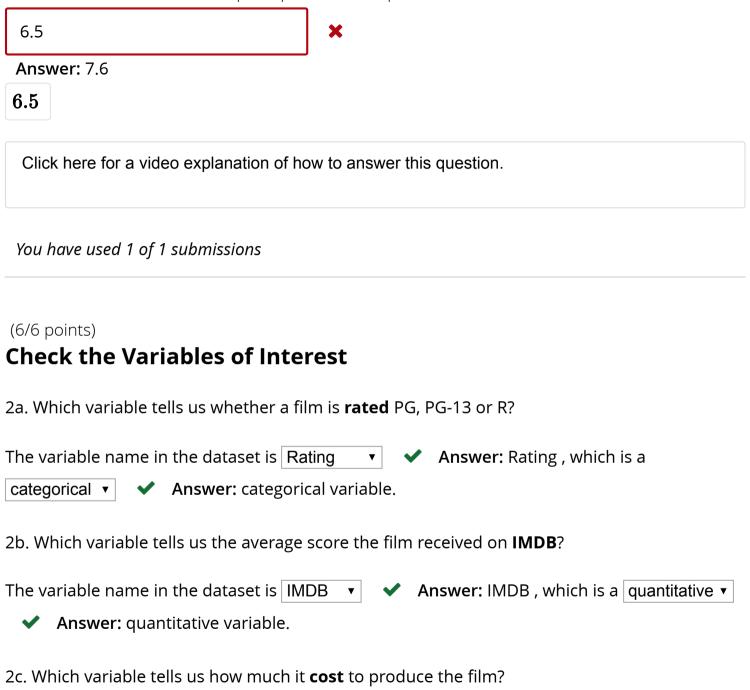
2

1b. What is the name of the highest ranked film made by Universal Studios? Double-check your spelling before you submit.



Answer: Jurassic Park

1c. What was the lowest IMDB rating given to a film that ranked in the top 10 grossing films of all time?



The variable name in the dataset is Budget Answer: Budget , which is a
quantitative ▼
Click here for a video explanation of how to answer this question.
You have used 1 of 1 submissions
(2/2 points) Reflect on the Method
Which method should we be using for this analysis and why?
3a. We will use ANOVA to help us answer each of these questions. Why?
 We want to determine if the category to which a film belongs has an impact on some other quantitative measure.
We believe the difference in scores given by film critics and IMDB users is significant.
 We cannot compare the means of these groups using any other method because the groups are related.

3b. We will conduct **post-hoc tests**, specifically Tukey's HSD, if the result of either ANOVA is significant. Why?

- We want to confirm that the assumptions of ANOVA have been met.
- We want to locate which group means are different from each other.
- We want to determine the average score for each category.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

















