



Bookmarks

- ▶ Important Pre-Course Survey
- ▶ Contact Us
- ▶ How To Navigate the Course
- ▶ Discussion Board
- ▶ Office Hours
- ▶ Week 1: Introduction to Data
- ▶ Week 2: Univariate Descriptive Statistics
- ▼ **Week 3: Bivariate Distributions**

Readings

Reading Check due
Mar 15, 2016 at 18:00
UTC

Lecture Videos

Comprehension Check
due Mar 15, 2016 at
18:00 UTC

R Tutorial Videos**Pre-Lab**

Pre-Lab due Mar 15,
2016 at 18:00 UTC

Lab

Week 3: Bivariate Distributions > Pre-Lab > Examine the Data



Bookmark

Reflect on the Question

Analyze the Data

Draw Conclusions

Lab 3: Professional Bull Riding

Over 1,200 bull riders from around the world are members of the Professional Bull Riders (PBR). They compete in more than 300 PBR-affiliated bull riding events per year. In the American tradition, the rider must stay atop the bucking bull for a full eight seconds. This data set includes information about the top-ranked bull riders for 2013. Rankings are based on a system which awards points for qualified rides at events throughout the season.

Primary Research Question

For the 2013 season, Is there a linear relationship between how often a rider placed in the Top 10 and the number of times he stayed on his bull for a full 8 seconds?

(3 points possible)

Check the Data

Let's begin by examining our data in R.

Lab due Mar 15, 2016
at 18:00 UTC

Problem Set

Problem Set due Mar
15, 2016 at 18:00 UTC

- ▶ Week 4:
Bivariate
Distributions
(Categorical
Data)
- ▶ Week 5: Linear
Functions

1. Open RStudio. Make sure you've installed the SDSFoundations **version 1.3** package.
2. Type `library(SDSFoundations)`. This will automatically load the data for the labs.
3. Type `bull <- BullRiders`. This will assign the data to your Workspace.
4. Look at the spreadsheet view of the data to answer the following questions.

Alternatively, you can use follow the steps in the "Importing a Data Frame" R tutorial video, and use the BullRiders.csv file. (Right-click and "Save As.") Make sure to **name** the dataframe "bull" when importing.

1. Open RStudio.
2. Click on "Import Dataset" button at the top of the workspace window. Choose *"from text file."*
3. Click on the location of the BullRiders.csv file you just downloaded.
4. Click on the BullRiders.csv file. Then, click Upload.
5. Look at the spreadsheet view of the data to answer the following questions.

1a. How many observations are in the dataset?

✗ Answer: 58

1b. How many of the first 10 riders in the dataset have been pro for 10 years or more?

✗ Answer: 3

1c. Of the top 15 riders so far in 2015, how many rides were completed by the rider with the fewest buck-outs in 2014?

✗ Answer: 0

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

(4/4 points)

Check the Variables of interest

Let's find the variables we need to answer the question.

Note: Be sure variable names are exact matches to the dataset. Fill-in-the-blank questions are case sensitive.

2a. Which variable tells us how many times the rider has placed in the Top 10 at the end of the 2013 season? The variable name in the dataset is:



Answer: Top10_13

2b. What type of variable is this?



Answer: Numerical

2c. Which variable tells us the number of times a rider stayed on his bull for the full 8 seconds in 2013? The variable name in the dataset is:



Answer: Rides13

2d. What type of variable is this?



Answer: Numerical

[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 the analysis and why?

3. We will use **correlation** to answer this lab question. Why?



We have two categorical variables that may be related.



We want to explore a linear relationship between two quantitative variables.

☐ We want to determine how a professional bull rider makes it onto the Top 10 list.

☐ We want to describe the distribution of a quantitative variable.

4. We should generate a **scatterplot** of these two variables before we continue our analysis. Why?

☐ We need to check the shape of the distribution.

☐ It is a good idea to make sure that we have quantitative data by plotting it.

☒ We want to confirm that the relationship is linear. ✓

☐ We need to identify how many cases are in the dataset.

[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.

POWERED BY
OPENedX

