Prepare for the Analysis

Primary Research Question

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

Breakdown Your Analysis

Let's break this analysis into its required steps:

1. Create a subset of the data which contains only those riders that have participated in at least one event in 2013.
2. Create a scatterplot of the two variables of interest.
3. Check to see that the relationship is linear. Plot a line of best fit as a guide.
4. If the relationship is linear, calculate the correlation coefficient.

5. Carefully examine any outliers.
6. Interpret what the correlation says about the linear relationship between these variables.

Here is the code you will use:

```
#Subset for riders that participated in at least one event in 2013
new_bull <- bull[bull$Events13  > 0 ,]

# Visualize and describe the first variable of interest
hist(new_bull$Rides13)
fivenum(new_bull$Rides13)
mean(new_bull$Rides13)
sd(new_bull$Rides13)

# Visualize and describe the second variable of interest
hist(new_bull$Top10_13)
fivenum(new_bull$Top10_13)
mean(new_bull$Top10_13)
sd(new_bull$Top10_13)

# Create a scatterplot
plot(new_bull$Rides13,new_bull$Top10_13)

# Add line of best fit
abline(lm(new_bull$Top10_13~new_bull$Rides13))

# Calculate the correlation coefficient
cor(new_bull$Rides13,new_bull$Top10_13)

# Create a correlation matrix
vars <- c("Top10_13", "Rides13")
cor(new_bull[,vars])
```

---

problem

1/1 point (graded)

1. Which cases will be selected by this line of code?

```
new_bull <- bull[bull$Events13 > 0 ,]
```

- riders that have completed zero Events in 2013
- riders that have completed at least zero Events in 2013
- riders that have completed more than zero Events in 2013 ✔

Submit You have used 1 of 1 attempt

problem

1/1 point (graded)

2. What will appear in the scatterplot produced by this line of code?

```
plot(new_bull$Rides13,new_bull$Top10_13)
```

- There will be a single data point for each bull rider. ✔
- There will be two data points for each rider (one for each variable).

Submit You have used 1 of 1 attempt

problem

1/1 point (graded)

3. Which value is **not** a possible output of the following line of code?

```
cor(new_bull$Rides13,new_bull$Top10_13)
```

- 0.75
- -1.02 ✔
- 0.61
- -0.04

Submit You have used 1 of 1 attempt

problem

1/1 point (graded)

4. A correlation matrix allows you to calculate multiple correlation coefficients at a time. Here, we are only asking for the correlation between Rides13 and Top10_13. If you wanted to include **other** variables as well, how would you do that?

```
# Create a correlation matrix
vars <-c("Top10_13", "Rides13")
cor(new_bull[,vars])
```

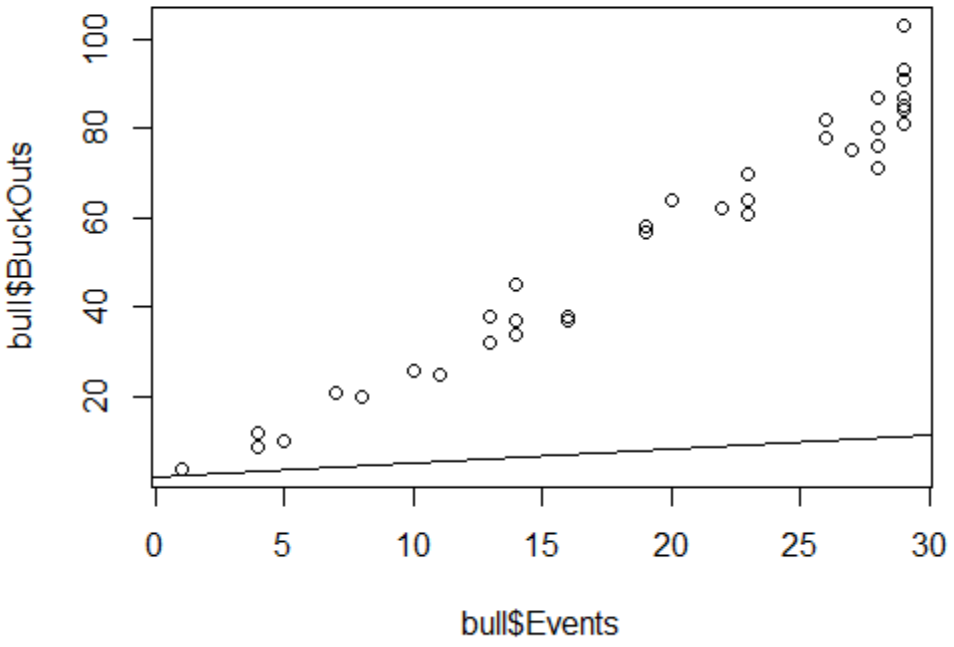
- Eliminate the line cor(new_bull[,vars]), because a correlation is only between two variables.

- (•) Add the variable names to the "vars" object. ✔
- ( ) You couldn't do that. You would have to run separate correlation matrices for each pair of variables.

Submit You have used 1 of 1 attempt

The following plot was produced using the code listed below:

```
bull<-BullRiders
plot(bull$Events12, bull$BuckOuts12)
abline(lm(bull$Events12~bull$BuckOuts12))
```



---

problem

1/1 point (graded)

5. In the above scatterplot, why does the line of best fit seem to not be going through the center of the scatterpot? (Refer to the code below and the dataset in R for help.)

- (•) The "Events" and "BuckOuts" variables should be switched in the abline command. ✔
- ( ) The "O"should not be capitalized in "BuckOuts".
- ( ) The plot was created with two variables that are categorical.
- ( ) The line of best fit goes through the plot accurately.

Submit You have used 1 of 1 attempt

© All Rights Reserved