Interpreting Correlations

Unit 2 - Lab 10

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

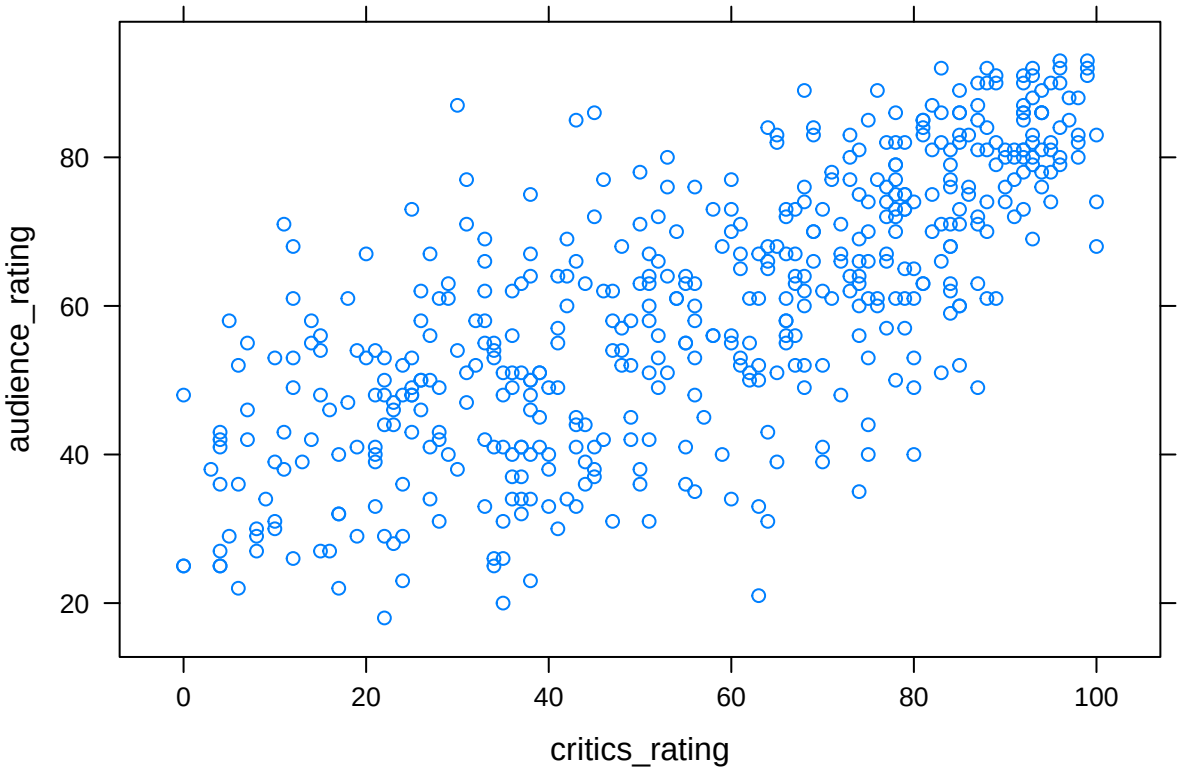
Some background...

- You will learn to measure the strength of a relationship between two numerical variables.
- For this lab, we will be using the `movie` data set to find correlation coefficients and interpret their appropriateness.

More background...

- Recall that a **correlation coefficient** describes the strength and direction of a LINEAR relationship between 2 *numerical* variables.
- Take a look at the variables in the `movie` data set.
- **List 3 pairs of numerical variables that you think could have an interesting relationship.**

Look at 2 of the variables in a scatter plot



Some questions

- **Does the plot show a positive or negative association?**
- **Is the association linear?**
- **What would you estimate the correlation coefficient to be? Do you think it should be positive or negative? Why?**
- But this doesn't seem too accurate...

So let's calculate it!

- We can use the `cor()` function to find the particular correlation coefficient of the previous two variables, `audience_rating` and `critics_rating`.

```
cor(audience_rating~critics_rating,
     data = movie)
```

- **What value was returned?**
- **How does this actual value compare with the one you estimated previously?**
- **Does this indicate a strong, weak, or moderate association? Why?**

More questions...

- **How would the scatter plot need to change in order for the correlation to be stronger?**
- **How would it need to change in order for the correlation to be weaker?**

What if we changed the data?

- **What happens to the correlation coefficient if you add a constant to one of the variables? Why do you think this is?**
- **Explain how you investigated this (include R code).**
- **What happens to the correlation coefficient if you multiple one of the variables by a constant? How do you explain this?**
- **Explain how you investigated this (include R code).**

Back to your own variables

- Select one of the pairs of variables you chose on Slide 3.
- **Plot the variables using the `xyplot()` function.**
- **Does the plot show a positive or negative association?**

- **Is the association linear?**
- **Calculate the correlation coefficient.**
- **What does this mean about the strength of the relationship between these two variables?**

Finishing up

- **Which of your pairs of variables has the strongest correlation? Report the variables and the corresponding correlation coefficient.**
- **Does it make sense that these variables would be linearly related? Why?**
- **Does it make sense that these variables would be strongly related? Why?**