If the line fits. . .

Unit 2 - Lab 11

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

Some background. . .

- In this lab, we will again be using the `movie` data set.
- But now, we're going to find a regression line between two numerical variables by
- Estimating a line of best fit by clicking 2 points on a scatter plot
- Calculating the line of best fit using the `lm()` function

Let's draw a plot!

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

How does it look?

- Recall that this plot shows a positive association between critics' ratings and audience ratings.
- **Which variable is represented by the x-axis? What does this mean?**
- **What can the x-variable tell you about the y-variable?**
- **If you were going to draw a line on the scatter plot to describe the relationship, how would you do so?**

We can estimate this line!

- Since we're not computers, we can't find the *exact* regression line just by hand.
- But we can still try!
- Once the scatter plot is created, we can run the `add_line()` function to add a line to our plot by selecting 2 points in the data.

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

```
add_line()
```

What happened?

- **What command appeared in your console once you typed `add_line()`?**
- **How did you decide on which two points to select?**
- **Once you selected your second point, what happened to the plot?**
- **What appeared in the console?**

We have an estimated equation for the line!

- **Rewrite the given equation with the correct variable names for x and y.**
- **Your neighbor's line is probably not exactly the same as yours. Discuss how you would determine which line was "best."**

Predicting from your line

- **Using the equation you found, calculate the predicted audience rating if the critics' rating is 85. How did you do this? Show your work.**
- **Find this value on your plot. Does it make sense based on the rest of the data points? Explain.**

We need a comparison

- The line you created earlier was just an estimate of the best line. We need to find the actual best line and compare our estimate to it.
- There is a linear model function (denoted as `lm()` in RStudio) that calculates the actual line of best fit between two numerical variables.

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

- **Record the output given in the console and rewrite it using the equation of a line.**

Interpreting what this equation means

- **Record the slope value from the equation. What does it mean in context of the variables?**
- **Record the y-intercept value from the equation. What does it mean in context of the variables?**

Predicting from the best line

- **Using the equation RStudio gave you, calculate the predicted audience rating if the critics' rating is 85. How did you do this? Show your work.**
- **How does this value compare with the value you found on Slide 8? Were they close or far apart?**
- **Do you think you did a good job of finding an estimated line of best fit? Why or why not?**