Regression (part 1)

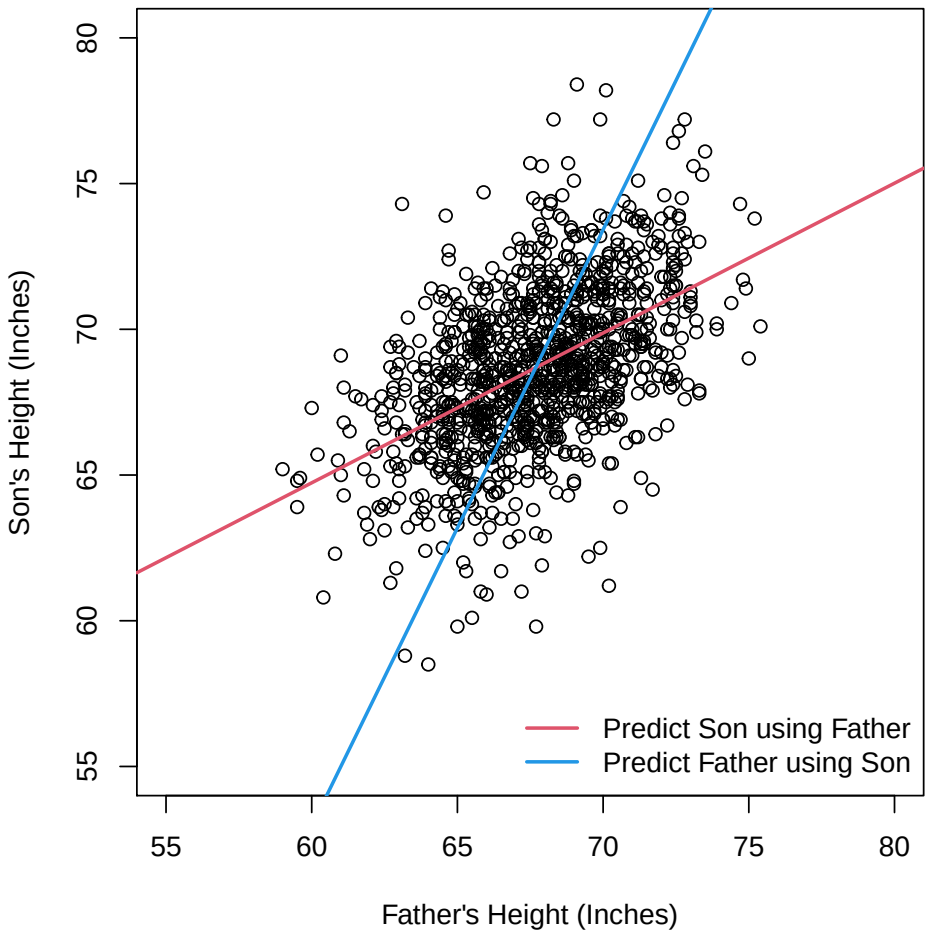
Ryan Miller

Introduction

- The past two presentations have introduced the correlation coefficient as a method for summarizing the relationship between two quantitative variables
  - Correlation is a **symmetric** statistical method: $r_{x,y} = r_{y,x}$, or it doesn't matter which variable is chosen to be "X" and which is chosen to be "Y"

Introduction

- Regression is an **asymmetric** statistical method: the choice of **explanatory** and **response** variables matters



Regression Lines

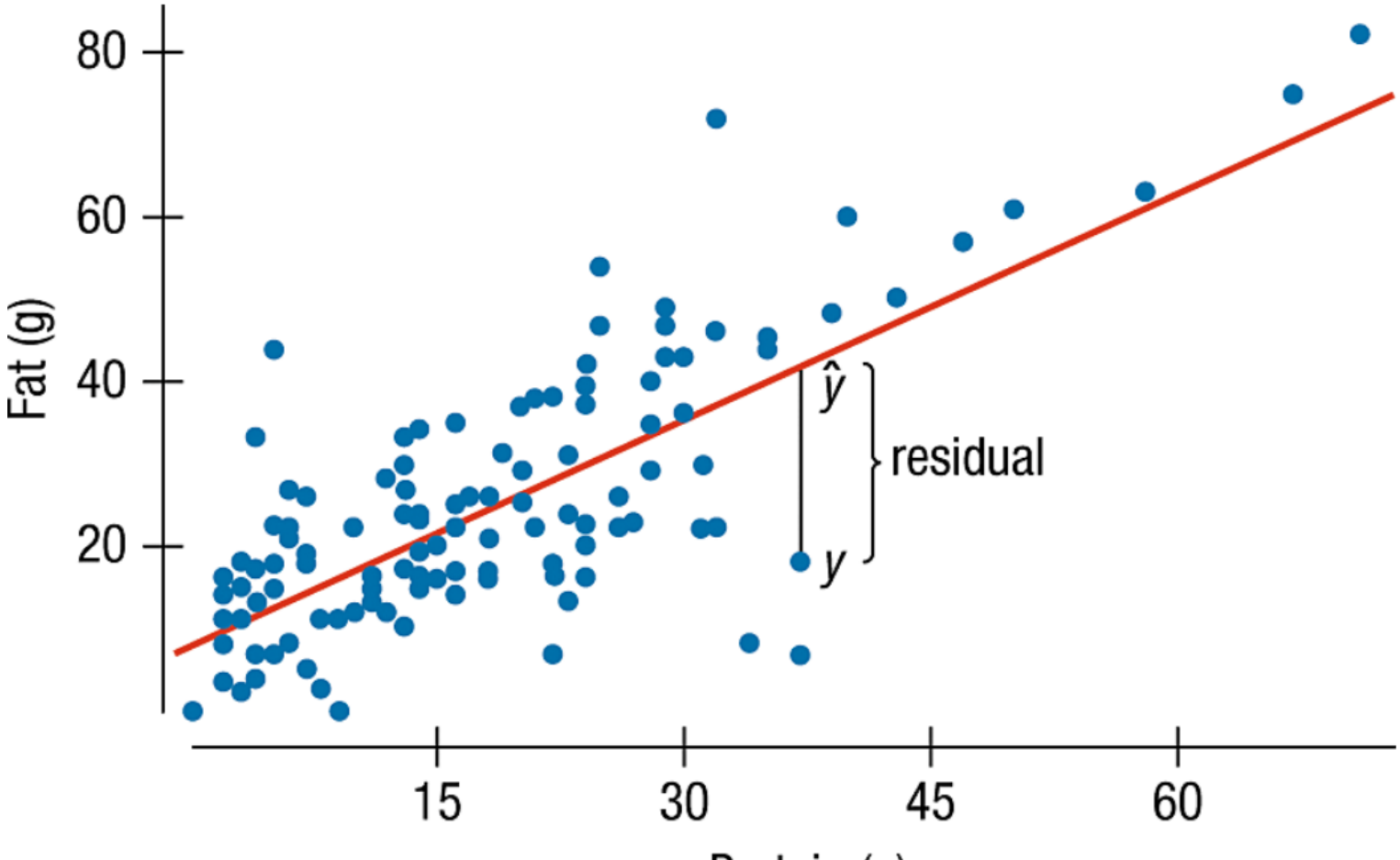
- Like any line, the regression line relating X and Y has two components, a **slope** and an **intercept**

$$\hat{Y} = b_0 + b_1 X$$

- in this notation, $\hat{Y}$ is the *predicted value* of the outcome variable
- X is the explanatory variable
- b_0 is the *estimated* intercept, or the predicted value when $X = 0$
- b_1 is the *estimated* slope, or predicted change in the outcome variable for a 1-unit increase in the explanatory variable

Regression Lines

- b_0 and b_1 are estimated from the data such that they minimize the squared **residuals**, or the distances between the predicted to observed outcomes
  - The example below relates Fat and Protein in the Burger King menu



Predictions

- The regression line can be used as a predictive tool:

$$\widehat{\text{Fat}} = 8.4 + 0.91 * \text{Protein}$$

- If we wanted an item with 20g of protein, we'd predict it to have $8.4 + 0.91 * 20 = 26.6$ grams of fat

Practice

1. Open the "Tips" dataset in the "data explorer" app
2. Use the "Summarize the Data" tab to find the slope and interecept of the regression line that uses "TotBill" to predict "Tip"
3. What does the *slope* of this line tell you about the relationship between Total Bill and Tip?
4. What does the *intercept* of this line tell you?
5. What does the regression line predict for a total bill of 20 dollars?

Practice (solution)

- $\widehat{\text{Tip}} = 0.92 + 0.11 * \text{Total Bill}$
- The slope of 0.11 suggests each one dollar increase in the bill leads to an 11 cent higher tip, or that people are tipping roughly 11% of the bill
- The intercept of 0.92 suggests the tip for a zero dollar total bill is 92 cents, or that people tend to tip at least 1 dollar even for very small purchases
- The predicted tip for a bill of 20 dollars is $0.92 + 0.11 * 20 = 3.12$ dollars