

# SDS 291 - Multiple Regression

January 27, 2020

## Lecture Notes

Here is some space to take notes on the slides we'll go through in class with some prompts for key ideas to pay attention to.

- **Sample, Population, Statistics, Parameters**

- **Types of Variables**

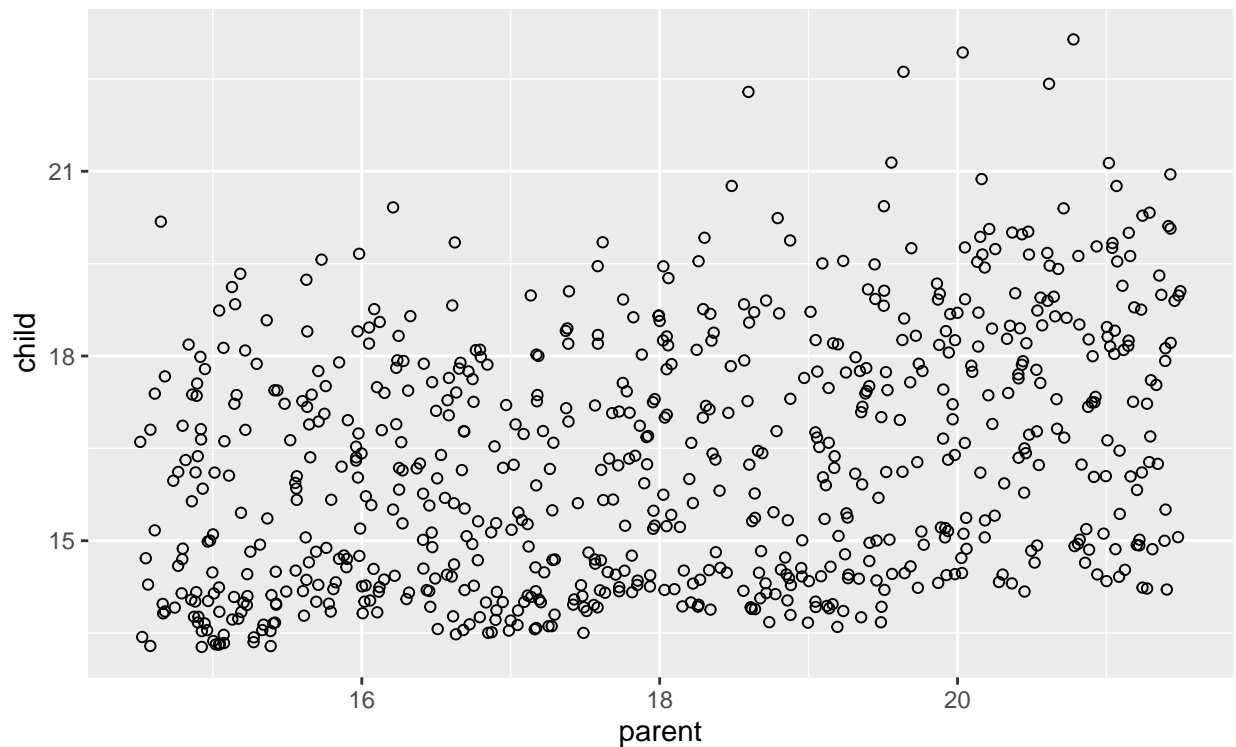
- **Describing a Scatterplot**

- **4 Step Process to Regression Modelling**

## Seed data Activity

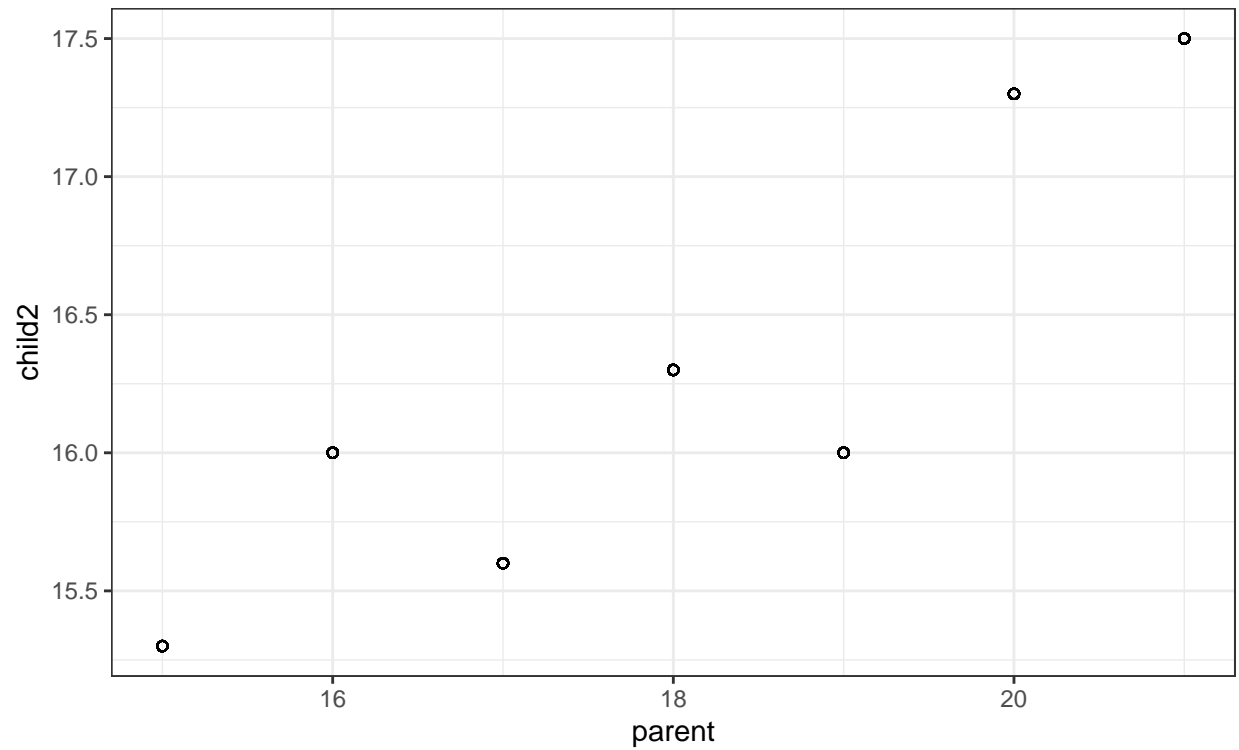
Your agricultural and biological scientist friend Frank works with seed for their pea plants (not just for work—they have lots of plants and identify as a #plantparent). They have child seed from parent plants and have been trying to anticipate how big the children seeds will be based on their parents' seed size. They want to know to what extent there is a quantitative relationship between parent and child seed size - measured in the diameter of the seed in millimeters (mm). Frank generated two scatterplots from a random sample of all of their data on child and parent seeds and needs your keen statistical eye and know-how to understand what they're looking at.

### 1. Scatterplot of Parent Seed Diameter (mm) and Child Seed Diameter (mm) - Version 1



1. Describe the scatterplot.
2. Draw a line that best fits these data.
3. Describe your reasoning to have drawn that particular line.

## 2. Scatterplot of Parent Seed Diameter (mm) and Child Seed Diameter (mm) - Version 2



1. Describe the scatterplot.
2. Draw a line that best fits these data.
3. Describe your reasoning to have drawn that particular line.

### 3. Numeric values

Frank also mentions that a statistician friend ran a regression model – you remember those from your intro statistics class! – but isn’t sure what it means. Can you help? They show you the following equation(s) that they have:

$$y = \beta_0 + \beta_1 Parent + \epsilon$$

$$y = 10.114 + 0.343Parent$$

1. In the second equation, what do each of the numbers mean? Please interpret each in a sentence for Frank.
1. Illustrate how to use this result to make predictions. What would this model predict a child seed’s diameter would be if the parent seed were 16, 17 and 20mm (make a prediction for each possible parent size)?

### 4. What to adjust for?

Frank knows that it isn’t just genetics that affect a child’s development but doesn’t know how to account for that quantitatively. You’re about to take a class on multiple regression and suggest that you might be able to “control for” some of the other things that might be associated with both child and parent seed size that could explain this relationship. Help brainstorm with Frank what some ideas are so they know what else to measure in their data other than just seed size.

1. List 2-3 other factors that are associated with child and parent seed diameter that could explain this relationship.