2.4: Exercises

Suppose you pick 50 random adults across the United States in January 2017 and measure how tall they are. For each of them, you also get accurate information about how tall their (biological) parents are. Now, using as your individuals these 50 adults and as the two variables their heights and the average of their parents' heights, make a sketch of what you think the resulting scatterplot would look like. Explain why you made the choice you did of one variable to be the explanatory and the other the response variable. Tell what are the shape, strength, and direction you see in this scatterplot, if it shows a deterministic or non-deterministic association, and why you think those conclusions would be true if you were to do this exercise with real data.

Is there any time or place other than right now in the United States where you think the data you would collect as above would result in a scatterplot that would look fairly different in some significant way? Explain!

It actually turns out that it is not true that the more a person works, the more they produce ... at least not always. Data on workers in a wide variety of industries show that working more hours produces more of that business's product for a while, but then after too many hours of work, keeping on working makes for almost no additional production.

Describe how you might collect data to investigate this relationship, by telling what individuals, population, sample, and variables you would use. Then, assuming the truth of the above statement about what other research in this area has found, make an example of a scatterplot that you think might result from your suggested data collection.

Make a scatterplot of the dataset consisting of the following pairs of measurements:

$$\{(8,16),(9,9),(10,4),(11,1),(12,0),(13,1),(14,4),(15,9),(16,16)\}\,. \tag{2.4.1}$$

You can do this quite easily by hand (there are only nine points!). Feel free to use an electronic device to make the plot for you, if you have one you know how to use, but copy the resulting picture into the homework you hand in, either by hand or cut-and-paste into an electronic version.

Describe the scatterplot, telling what are the shape, strength, and direction. What do you think would be the correlation coefficient of this dataset? As always, explain all of your reasoning!