

Richard Howell

Weekly paper 4

The first of the main idea I got from chapter 4 is how many different things a missing value might mean. It could mean that there was a data entry error, that the data was purposefully left out, or that the data being left out serves some purpose like having the value of false. The second main idea that I got out of this chapter was how useful sampling can be. Sampling, or using a part of the data to represent the whole, can give us a quick summary of the data as long as outliers are not represented more than they are in the original data. The third idea that I got most heavily from the chapter was how the quality of data relies mostly on us. We decide how missing values are handled, and we decide how sampling occurs, so we decide how much information the data actually gives us.