

We have included data files with each question that are required to answer them. All data has been simulated. Please read each question carefully and take note of what we are asking you to submit for each question.

Question 1

The fish processing plant is interested in optimizing the yield of their plant and thanks to the latest data dump you are in a position to help! Create a model to predict the yield based on the data we have provided you. Please submit your predictions on the test dataset, as well as a paragraph explaining why your model is the best choice, and your code.

Files: *Q1TrainData.csv* *Q1TestData.csv*

Question 2

Data from a new fish processing plant in Ecuador has just hit your desk. You begin your preliminary explorations of the data and discover that the yield values follow this distribution (Fig 1). Analyze this data and try to uncover what you can learn about the data with regards to this distribution. Please submit a paragraph explaining what you have learned about the data, a paragraph detailing and justifying your choice of methods, your code, and optionally any graphs that would aid us in understanding your findings.

Files: *Q2Data.csv*

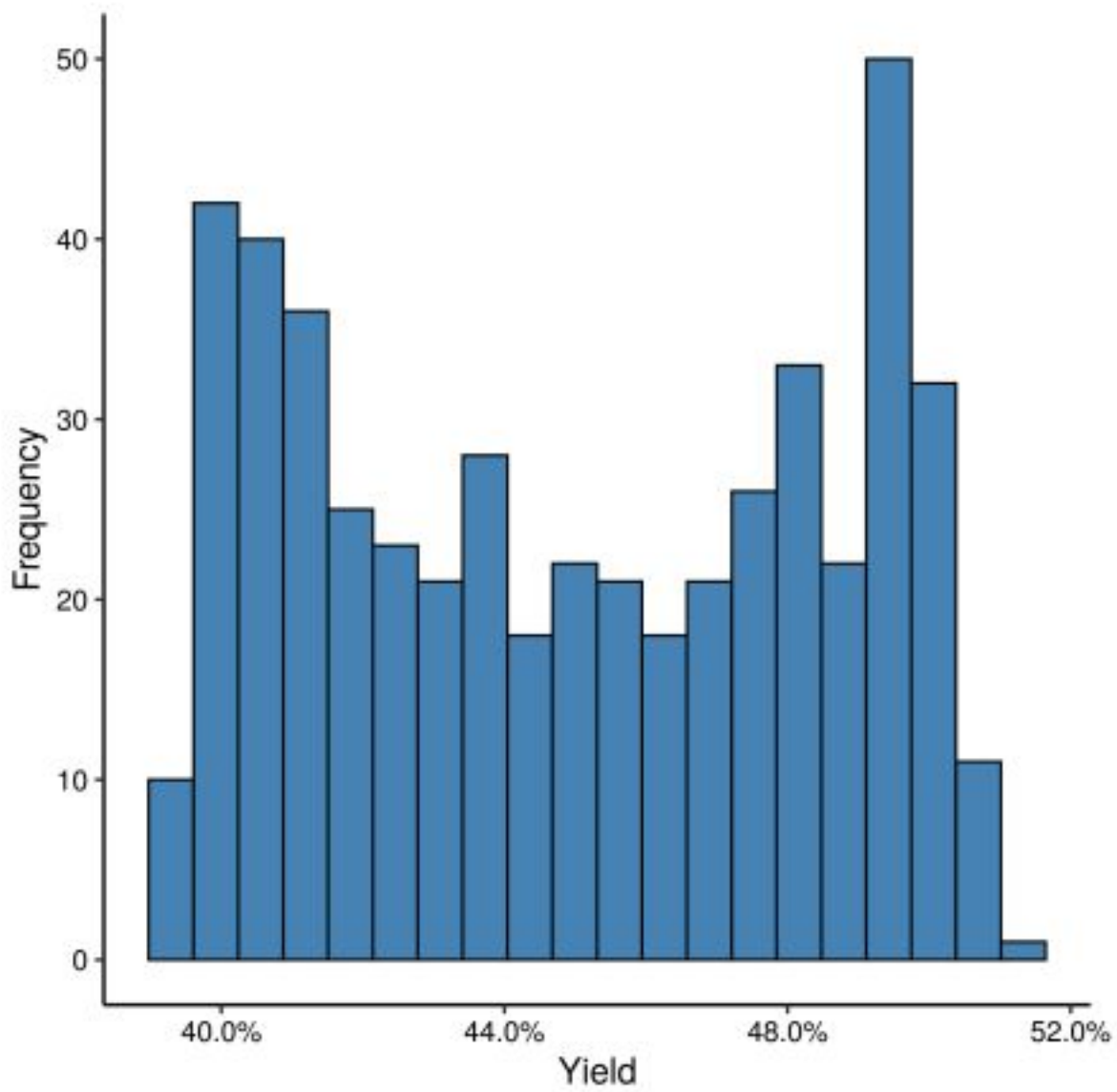


Figure 1: Yield values.