Assignment 64-4

Riley Paddock

November 2020

Part A

In General we would expect the RSS for the Cubic regression to be lower because it can fit the data more specifically. We would not be able to tell which one fits the testing data better but for the simple training data the Cubic regression would have a lower RSS.

Part B

The linear regression would have a lower RSS on the testing data because the X and Y are linearly related. Meaning that the linear regression would be closer to those data points in the testing set. Whereas the cubic would just be specific to the data in the training set.

Part C

If the data is not linear the cubic regression will fit better to the training data than before. Because there is more variablity in the training data the linear regression will have a harder time fitting to it and will have a higher RSS where the cubic has the ability to curve and fit more accuratley to the data so it will have a lower RSS.

Part D

If we do not know the circumstances of the relationship between X and Y we cannot know whether the linear or cubic regrression will have a lower RSS in relation to the testing data. With the training data we can assume the higher degrees will fit better no matter the data but with the testing data it depends on the true relation of X and Y.