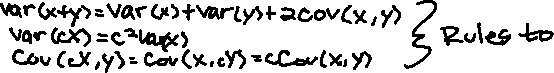
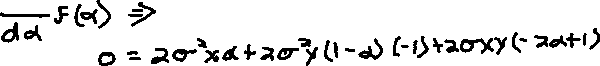
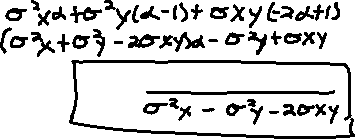
Alexa Summers

Gireesh Kumar Muppalla

Santhoshini Sree Bolisetty

Homework #3—CS 5565

1.­



2a. The jth observation has a probability of 1/n of being the first bootstrap sample, so the probability that it is not the first bootstrap sample is 1-1/n.

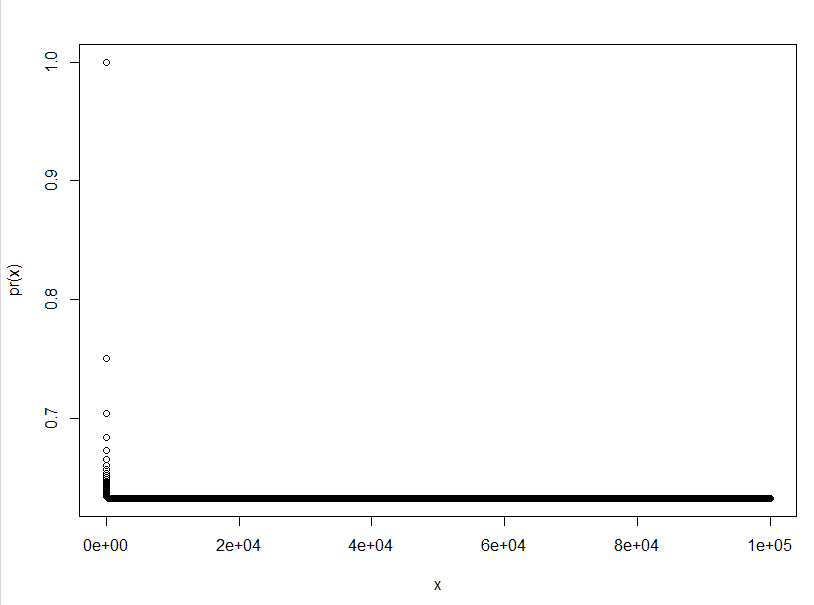
2b. The jth observation has a probability of 1/n of being the second bootstrap sample, so the probability that it is not the second bootstrap sample is 1-1/n.

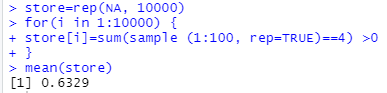
2c. (1-1/n)(1-1/n)(1-1/n)… = (1-1/n)n. Each observation has an independent chance of equaling the jth, so after applying the product rule, we end up with (1-1/n)n.

2d. n = 5 -> 1−(1−1/5)5=1−(4/5)5=67.2%

2e. n = 10 ->1−(1−1/100)10=1−(99/100)100=63.4%

2f. n = 10,000 -> 1−(1−1/10000)10000=63.2%

2g.

2h.

The asymptote is 63.29% and is approached rapidly. The selection probability is almost equal to the answer for 2f (n = 10,000 -> 1−(1−1/10000)10000=63.2%), as expected.

3a. Take a set of observations and split them into non-overlapping groups (k). These groups will act as the remainder of a testing set. When the resulting MSE estimates are averaged together, the test error can be estimated.

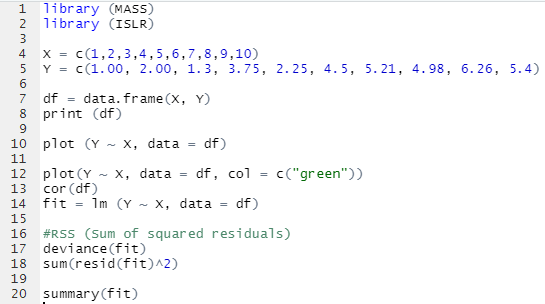
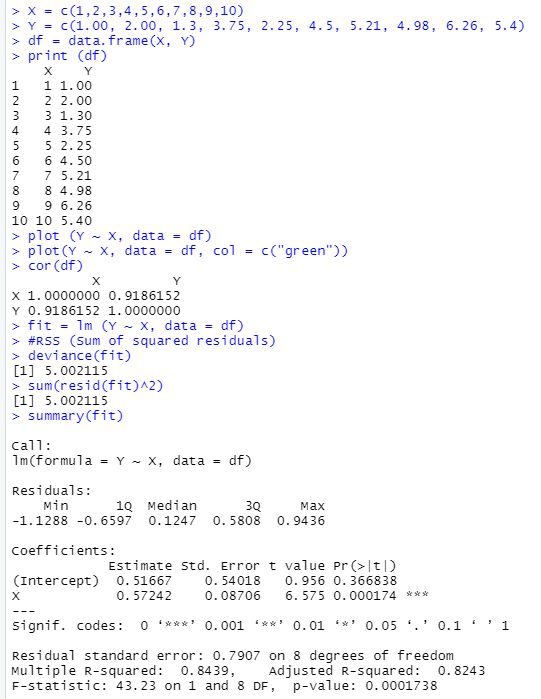
3bi. Advantages: easy to implement and relatively simple to understand.

Disadvantages: The estimation of the test error can have a high variation depending on observations in the training and testing sets. Only a subset of observations are used to fit the model, which tends to result in a worse performance.

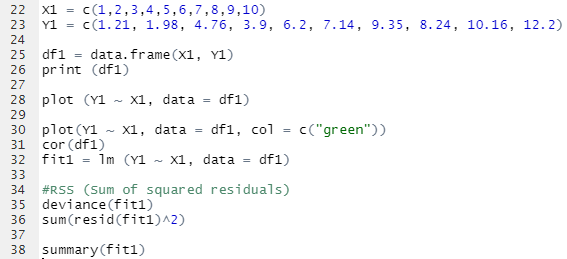
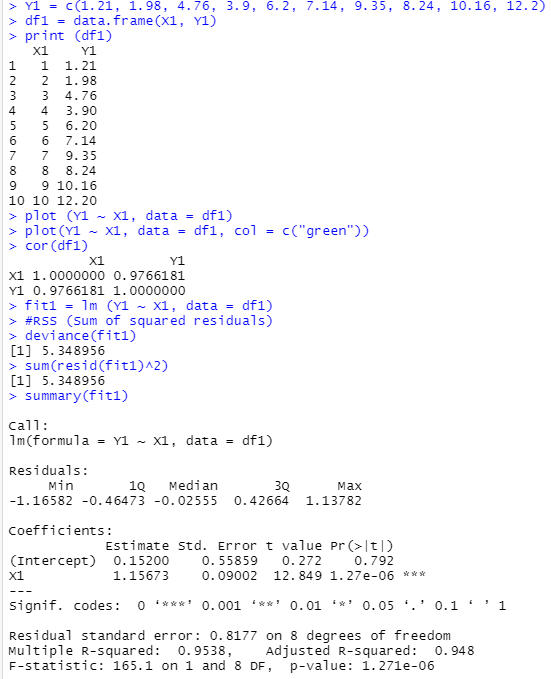
3bii. Advantages: Less bias, less variable MSE.

Disadvantages: Computationally intensive—takes a long time.

4. We could use a bootstrap distribution by sampling observations from the original dataset, and then fitting a new model for each repetition, and then looking at the RMSE of all the estimates.

5.

Coefficient of B̂0: 0.51667Coefficient of B̂1: 0.57242  
RSS: 5.002115  
RSE: 0.7907  
R2: 0.8439  
T-statistic: 6.575  
P-Value: .000174  
  
Reject the null hypothesis.

6.

Coefficient of B̂0: 0.15200Coefficient of B̂1: 1.15673  
RSS: 5.348956  
RSE: .8177  
R2: 0.9538  
T-statistic: 12.849  
P-Value: 1.27e-06  
  
Reject the null hypothesis.