1. Is *lg(N)* or *N* a better approximation for your measured heights? Why?

   Since O(lg(2^20)) is faster than O(2^20), lg(n) is a better approximation for measured heights.

2. How much variance is there in the height of these trees? Is this surprising? Why or why not?

   There is not significant variation in the height of the trees since the standard deviation is about 2.7, so it is not surprising.

3. How many duplicate values were there (on average)? Is this surprising? Why or why not?

   The average of the duplicate values is 128.6. Due to there is not significant difference in the duplicate values, it is not surprising.

4. The "big Oh" notation *O(f(n))* implies that there are constants *a* and *b*, such that *a * f(n) + b* describes the reality for which *O(f(n))* is an approximation. A balanced BST of *N* items has height *O(lg(N))*. Theoretically, there should be values *a* and *b* such that *a * lg(N) + b* describes the actual height of a BST built using random data. Using your experimental results, what is a reasonable value for *a*?

   Based on the result, the average height = *a * log(N)* is 50.4 = *a * log(2^20). a = 50.4/log(2^20) = 3.63559…*