Question 1

The following experiment is conducted. We randomly select tree in a forest, measure its height and record it as x_1. Then we again select random tree in the same forest, independently on the first choice, measure its height and record it as x_2, and so on.

1. Can we consider $(x_1, \ldots, x_n)$ as an i.i.d. (independent and identically distributed) sample from some random variable? I.e. can we assume that these values are obtained as independent realizations of some random variable X?

2. Now we sort values $(x_1, \ldots, x_n)$ in ascending order and thus obtain new sequence of numbers denoted by $(y_1, \ldots, y_n)$. Can we say that $(y_1, \ldots, y_n)$ is an i.i.d. sample from some random variable?

Hint: it can be helpful to solve the following problem before answering question 2. Let X_1 and X_2 be two independent Bernoulli random variables with $p = 1/2$ (i.e. they take values 0 and 1 with equal probabilities.) Consider new random variables $Y_1 = \min(X_1, X_2)$ and $Y_2 = \max(X_1, X_2)$. Is it true that they Y_1 and Y_2 are independent random variables? Use definition of independent random variables.

Explain your answers.

1. We can consider $(x_1, \ldots, x_n)$ as an i.i.d. sample from some random variable X because the heights of the trees are measured randomly and independently of each other.

The height of one selected tree does not affect the height of another selected tree. Therefore, each height measurement can be considered as an independent realization of a random variable X. Here we should make an assumption, that after having been selected the tree is "returned" back to the sample space and has a chance to be selected again. Otherwise, the probability of selecting each next tree will depend on the probability of the previously selected trees.

As to the identical distribution, the selection process is done randomly and uniformly for each tree, and the height measurement is performed only once per each selection, which implies that the distribution of X is identical and even equiprobable for all the trees.

Hence, the measurements are independent and identically distributed, and we can treat them as a random sample from X.

2. We cannot consider the sorted sequence $(y_1, \ldots, y_n)$ as an i.i.d. sample from some random variable due to the loss of independence caused by sorting.

Rearranging the original measurement values alters their random order and creates dependencies between them, as now the order of some measurements can affect the probability of other measurements. For instance, if y_1 is significantly small, we know that the rest of the measurements must be larger than y_1.

Such breach of independence between the measurements, makes it impossible to consider $(y_1, \ldots, y_n)$ as independent and identically distributed, and consequently, it cannot be treated as a random variable sample.