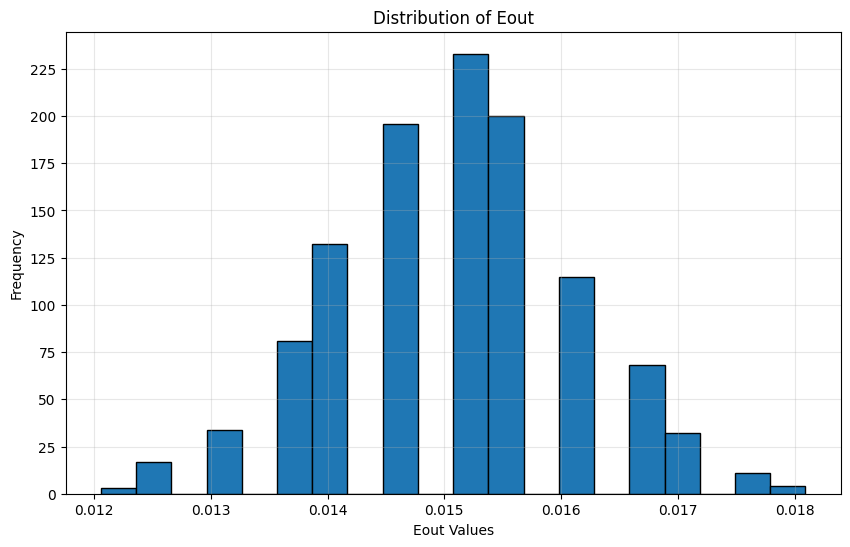
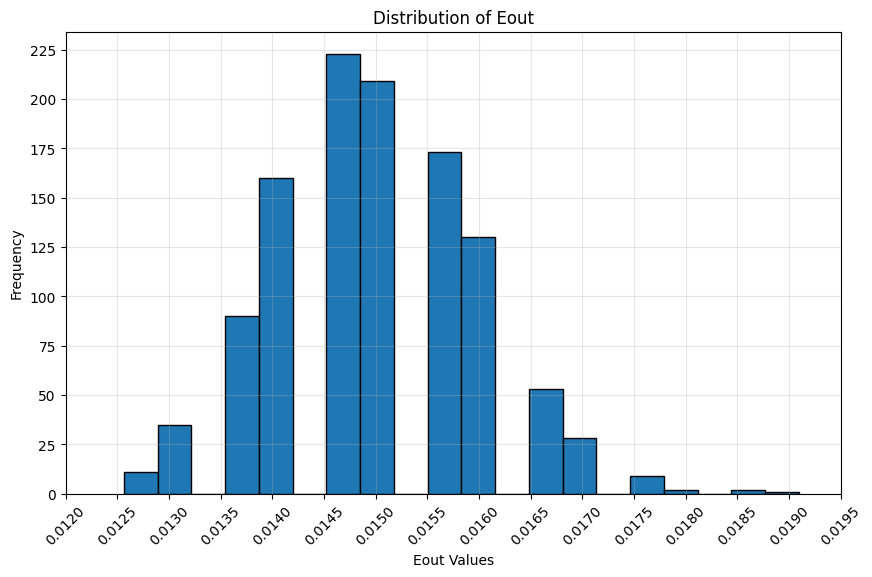
# HTML homework 5: q12 report

The resulting picture is as follows:

(q12)

To find the difference, the plot below is the result of the previous problem:

(q11)

From the two plots we can see that the subtrain / validation split is more dispersed, and the variability of the 3-fold cross validation is smaller.

The reason why the subtrain / validation split is more dispersed may due to the fact that the validation set size is not that big compared to the 3-fold version. In the previous question, the validation set size is 3876, and in this question, using 3-fold splits the training set into 3 sets, with each about 3959 examples, which is slightly higher. Also, is calculated by the mean error over using each of the 3 sets as the validation set, therefore these 2 reasons introduce stability in the 3-fold version.

On the other hand, the subtrain / validation split relies only on one split, if the validation set is not that representative, our selection of may be biased, this would magnify the effect of the unstableness in the previous question.

## Code:

The markdown screenshot below describes the modification of this problem, other parts like reading in the data and the error function…, are the same as problem 10: