③ 1) Bootstraping & It is a method of sample reuse. The idea is to use the observed sample to estimate the population distribution. Then samples can be drawn from the estimated population and the sampling distribution of any type of estimator can itself be estimated.

There are three types of bootstraping &

i) Nonparametric & A sample of same size as the data is taken from the data with replacement. what does this mean It means that if you measure 10 samples, you create a new sample of size 10 by replicating some of the sample that you have already seen and omiting others.

ii) Semiparametric & It can only reproduce the items that were in original sample. It assume that the population includes other items that are similar to observed sample by sampling from smoothed version of the sample histogram.

iii) Parametric & It assumes that data comes from a known distribution with unknown parameters. You estimate the parameters from the data that you have and then you use the estimated distributions to simulate the samples.

All above samples methods are simulation based ideas.

Q. Cross Validation & It is a method that uses the same data to both train the model and obtain a less biased estimate of prediction error than the direct estimate. The basic idea is to split the training data into two subsets - one is used to train the prediction rule and then the other subset is used to assess prediction error. To use the data efficiently, this is repeated with multiple splits of the data.

There is a small problem with this method for assessing prediction error. The final predictor will be based on N, but estimated prediction error is based on predictor developed on smaller sample $N > N - N/k$. So Cross-validated estimate of prediction error might actually be pessimistic - might have slightly better prediction error than you think. However with 10-folds cross-validation can't be too far off because you are using at least 90% of your sample.