Definition of Subsampling: Suppose we have a sample

$$\mathbf{Y} = Y_1, \ldots, Y_n \text{ from a distribution } F$$

Subsampling is drawing smaller samples or subsamples of size $m < n$ from $\mathbf{Y}$ without replacement in order to draw statistical conclusions about any parameter of interest.
The required statistic is calculated from each sample and these values are used to construct an approximation of the appropriate sampling distribution.

Advantages of Subsampling over ordinary bootstrap

Subsampling achieves more accuracy in estimation than ordinary bootstrap under less stringent regularity conditions than those required by bootstrap.
It provides asymptotic consistency even under extreme weak conditions and is hence a popular non-parametric tool.
Some of its advantages over bootstrap are given below.

1. Subsampling has a distinct advantage over bootstrapping when the sample available is from time series data.
   Time series is dependent data, i.e., the ordering of data according to time is extremely important. It makes little or no sense in drawing identical and independent resamples using bootstrap technique from time series data.
   Using sub-sampling technique, we can obtain samples and yet maintain the dependence according to time. The statistics obtained may not be the unbiased estimates required, but they can be easily amended to obtain the required estimates.
   For example, if $(Y_1, \ldots, Y_n)$ is a time series, the subsequences $(Y_s, \ldots, Y_{s+m-1}),\ s = 1, \ldots, S,\ S = n - m + 1$ are the subsamples.

2. When constructing confidence region for parameters.
   Suppose we have sample

   $$\mathbf{Y} = Y_1, \ldots, Y_n \text{ from an unknown distribution } F$$

   and $\theta \equiv \theta(F)$ is the parameter for which we want a confidence region.
   The confidence region is constructed from a statistic $\hat{\theta}_n$ that converges weakly to θ at a rate τ_n.(A common choice of τ_n is $\sqrt{n}$).
   Let $\hat{\sigma}_n$ be an estimator of σ, σ^2 being the asymptotic variance of $\tau_n\hat{\theta}_n$.
   Suppose $J_n(F)$ be the sampling distribution of $\frac{\tau_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n}$. Let $J(F)$ be a non-degenerate distribution such that

   $$J_n(x,\, F) \to J(x,\, F)\ as\ n \to \infty$$

   We draw $S = \binom{n}{m}$ sub-samples each of size m and let $\hat{\theta}_{n,m,s}$ and $\hat{\sigma}_{n,m,s}$ be the estimates of $\hat{\theta}$ and $\hat{\sigma}$ obtained from the sth sub-sample. The sub-sampling distribution of $\frac{\tau_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n}$ is

   $$L_{n,m}(x) = \frac{\sum_{s=1}^{S} I[\frac{\tau_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n} \le x]}{S}$$

   If $m \to \infty$ and $max(\frac{m}{n},\ \frac{\tau_m}{\tau_n}) \to \infty$ as $n \to \infty$ then

   $$sup_x[L_{n,m}(x) - J(x, F)] = o_p(1)$$

   Hence quantiles of $L_{n,m}$ can be used to construct confidence sets of θ with asymptotically correct coverage.

Bootstrap approximations can provide analogous results, but under more strong assumptions. For example, the convergence of $J_n(F)$ to $J(F)$ must be locally uniform in F . Nonuniformity in convergence is responsible for bootstrap failure.

3. Subsampling can be used to remedy bootstrap.If subsampling is done with replacement, we get *m out of n* bootstrap .
   With assumption $m/n \to 0$ and additional assumption $m^2/n \to 0$, sub-sampling and *m out of n* bootstrap give similar asymptotically valid conclusions.

Disadvantages of Subsampling

1. The proper choice of the sub-sample size is difficult. If the choice is not perfect, high-order accuracy cannot be obtained from sub-sampling.

2. When bootstrap is valid, it is usually preferred to sub-sampling. When *m out of n* bootstrap is valid, it is also preferred to sub-sampling.

But if validity of bootstrap cannot be verified or under minimal conditions or complicated data strucutres, sub-sampling is always the preferred solution.