

Linear Regression Boot Strap

Objective:

- Reasons for using Bootstrap for linear Models
- How to generate the Bootstrap sampling distribution

Bootstrap Motivation :

- If the distributional assumptions like Normality are not met then we cannot perform hypothesis testing or get valid hypothesis testing.
- If the distributional assumptions are not known then we use the non parametric Boot Strap which gives valid hypothesis tests, confidence intervals and standard error.
- The only assumption that we are taking is that the sample that we have is representative of the population.
- Bootstrapping is a resampling method which is more computationally intensive as compared to traditional inference.

Bootstrap Process:

Objective: We wish to estimate a population parameter θ (like the slope coefficient) given an estimate $\hat{\theta}$.

Steps:

1. We have a representative sample of size n from the population
2. We now resample with replacement from the sample to create a bootstrap sample. The boot strap sample size could be of the same size n .
3. Create k such bootstrap samples
4. For each of the k bootstrap samples estimate the $\hat{\theta}_1^*, \dots, \hat{\theta}_k^*$
5. Using these estimate generate an empirical sampling distribution for $\{\hat{\theta}_1^*, \dots, \hat{\theta}_k^*\}$ and then estimate the properties of sampling distribution of $\hat{\theta}$.

<http://www.statmethods.net/advstats/bootstrapping.html>

Rcode