



Day 20 of #100daysofmathandstats: Data sampling Concepts(Contd...)

By Harsh Kathiriya



Outline

- Bootstrap sample
- Resampling
- Resampling vs bootstrapping

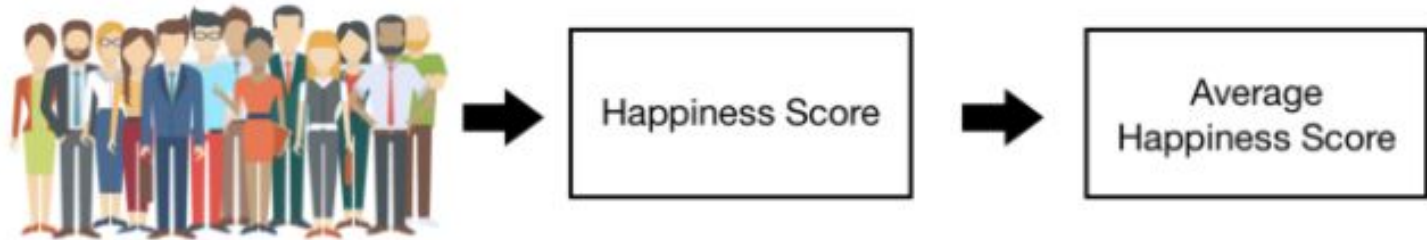
Bootstrap sample



- A sample taken with replacement from an observed data set.
- A bootstrap sample is a smaller sample that is “bootstrapped” from a larger sample. Bootstrapping is a type of resampling where large numbers of smaller samples of the same size are repeatedly drawn, with replacement, from a single original sample.
-

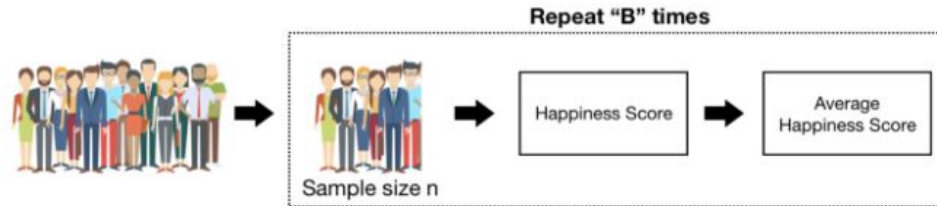
Example of bootstrapping

- Lets say your company is performing poorly. You think this is correlated to happiness. You want to ask your employees to rate their happiness on a scale of 1–10.



Example of bootstrapping (Contd...)

- If you're a small start-up of 15 people, seems doable. But what if you're a large corporation of 1000s of employees? No one has time to interview that many people. Instead of speaking to each employee, we can find the happiness scores of a random sample of n employees. If we do this B times, we can find the happiness scores of $n \times B$ employees, and use this as an estimate of "happiness" across the company.



Resampling



- The process of taking repeated samples from observed data; includes both bootstrap and permutation (shuffling) procedures.
- Resampling is the method that consists of drawing repeated samples from the original data samples. The method of Resampling is a nonparametric method of statistical inference.
- In other words, the method of resampling does not involve the utilization of the generic distribution tables (for example, normal distribution tables) in order to compute approximate p probability values.

Resampling vs bootstrapping



- Sometimes the term resampling is used synonymously with the term bootstrapping, as just outlined.
- More often, the term resampling also includes permutation procedures ,where multiple samples are combined and the sampling may be done without replacement.
- In any case, the term bootstrap always implies sampling with replacement from an observed data set.



Thank you

Github Link: <https://github.com/harsh9898/100daysofstatandmath>

Don't forget to post your queries or feedbacks on the post.

Share or like for the benefit of others.