

## Activity - Proportional Stratified Sampling

Download the file [startup\\_funding](#)

### Problem Statement :

We want to find out the maximum amount of funding given to any startup in the following regions -

Bangalore

NCR (which includes New Delhi, Gurgaon and Noida)

Mumbai

Pune

Hyderabad

So create your population data first by extracting those startups which are in given cities.

In the last question, we have used "Stratified Sampling" to sample our data. But you must have realized that our data is having different proportions of startups which are in given cities. So randomly selecting 20 values from each stratum is not a good idea. The size of different strata is quite different.

Ideally, we should take a number of samples from each stratum depending on their proportion to the complete dataset.

So here we'll apply "Proportional Stratified sampling". To use this, first, you need to calculate the number of data points that we need to take from each stratum and that can be calculated as -

$$(\text{No\_of\_values\_ith Strata} / \text{population\_size}) * \text{Sample\_Size}$$

Then take a respective number of values from each stratum based on its proportion (using SRS) and create your sample.

Find and print Sampling errors.

### Note:

Take the city name "Delhi" as "New Delhi".

Check the case-sensitiveness of cities also. That means - at someplace, instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore"