README:

The following code is a robust and efficient python library used to generate ranked set sample datasets. Ranked set sampling is an alternative to simple random sampling and is particularly useful when performing statistical analyses on smaller sample sizes. Traditional ranked set sampling consists of the following parameters:

* ‘m’= the number of units in each row subset (usually is optimized at m=6)
* ‘r’ = the number of cycles to generated the ranked set sample. This can be found by taking

*r=[n/m]* where ‘n’ is the sample size of the initial dataset.

In order to generate a ranked set dataset, the sets the ‘MainPythonScript.py’ as the startup project. From there, simply enter the path for the initial dataset as a .csv in order to import the data as a pandas dataframe. Executing the program will output a ranked set sample subset based on the indices of the initial dataset. It is also possible to execute the number of ‘resamples’ (i.e. bootstrap) if the user desires multiple ranked set datasets.

Project dependencies:

-numpy

-pandas

The python code contains 5 classes to generate a ranked set sample dataset:

-CyclesArrayGenerator= used to generate the whole array that will be parsed based on the number of cybles ('r') and the number of units in each set ('m') . The array is 1-dimensional but will eventually be parsed as an (*m x r)* 2D array.

-MainRSSamplingClass= the main entry point of the program after all the parameters have been set including the initial dataset, the number of units in each set, and the number of cycles. This class will return the result of one ranked set sample as a panda's dataframe based on the initial dataset. There is also an extra parameter if the user wishes to run the script from R (python code can easily be run from R using the reticulate package).

-RSSRowSorter= Once the 2d array has been created and parsed properly, this class will go through the rows and sort all the indices.

-RSS2DArrayGenerator= Parses the giant 1-dimensional array into an (*m x r)* 2D array. The class then gets the index values at the diagonals (shown below). These diagonal indices are then mapped back to the row values in the initial dataframe.

A picture containing hanging, dog

Description automatically generated

*References:*

*-McIntyre, G.A. (1952) A Method for Unbiased Selective Sampling Using Ranked Sets. Australian Journal of Agricultural Research, 3, 385-390.*[*http://dx.doi.org/10.1071/AR9520385*](http://dx.doi.org/10.1071/AR9520385)

*-*