Subject: Combining individuals DHS datasets in Python
Posted by Triphon on Thu, 25 Jul 2019 12:22:54 GMT
View Forum Message <> Reply to Message

Hi,

I am merging MR and IR datasets with HIV status (AR) for multiple countries. From what I've read, weights are hiv05 (preferred over v005/mv005), cluster (id) is v021/mv021 and stratum depends but let's say v023/mv023.
I am coding in Python (for various reasons) and there is no package/library such as svydesign in R. I could use R and then back to Python but switching from pandas categorical to R factor is not straightforward (data type not string). In Python, I can handle the weights, but not the cluster and the stratum (to my knowledge). Does someone knows what would be the best option here? For prediction and variables selection purposes, not taking into account cluster and stratum would be problematic?

Subject: Re: Combining individuals DHS datasets in Python
Posted by Mamadou S DIallo on Thu, 31 Dec 2020 21:43:01 GMT
View Forum Message <> Reply to Message

Dear Triphon

I came across this post. I know it is old but in case you are still interested in using Python to analyze survey samples, samplics may be useful for you.

samplics is a Python package that I have been developing. It is not yet feature-complete as survey R but the pieces included so far have been extensively tested. I am working on adding needed features. Feedback from users will be extremely valuable to guide the next items to add to the package. If you are still analyzing survey data, I will be very interested in your feedback. You will find samplics' documentation at https://samplics.readthedocs.io/en/latest/. If you are interested, do not hesitate to send me an email at msdiallo@quantifyafrica.org.

Happy new year and all the best for 2021.