

---

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?

---