

---

Subject: Re: Pooling male and female files

Posted by [Bridgette-DHS](#) on Mon, 02 Oct 2017 18:40:27 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

Regarding your second question, we do not have the sampling fractions or weights for households within clusters. The weight we provide is for the combination of sampling clusters and sampling households within clusters. There have been several earlier posts on this.

Regarding the first question, if you open the PR file and enter "tab1 hv117 hv118" you will see that this survey subsampled only half of the men. When you combine with women, you need to approximately double the weight for the men.

The following routine will calculate the correct weight for the men relative to the women

```
use e:\DHS\DHS_data\PR_files\NGPR6AFL.dta, clear
```

```
* Reduce the PR file to men who are eligible by age and are de facto
```

```
keep if hv105>=15 & hv105<=49
```

```
keep if hv103==1
```

```
keep if hv104==1
```

```
* Total hh weight for men who are eligible by age and are de facto
```

```
summarize hv005
```

```
scalar W=r(sum)
```

```
* Total hh weight for men who are eligible by age and are de facto and are subsampled
```

```
summarize hv005 if hv104==1 & hv118==1
```

```
scalar W1=r(sum)
```

```
* Calculate the ratio
```

```
scalar factor=W/W1
```

```
scalar list factor
```

```
use e:\DHS\DHS_data\MR_files\NGMR6AFL.dta, clear
```

```
* Multiply mv005 by the ratio and round to an integer
```

```
gen mv005_rewtd=round(mv005*factor)
```

```
* Then append the IR and MR files and use mv005_rewtd as the weight for men
```

It would be slightly better to calculate separate factors within each stratum. You can re-normalize so that the mean weight in the file of women and men is 1000000, but if you will only be using pweights, you can skip that step because Stata (with pweight) always normalizes the weights to have a mean of 1.

