

---

Subject: Re: Côte d'Ivoire 1998 DHS  
Posted by [Bridgette-DHS](#) on Fri, 02 Sep 2016 10:30:37 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

Here's how to do this in Stata:

```
* open the PR file  
use ...CIPR3AFL.dta, clear
```

```
set more off
```

```
* reduce to one record per household  
keep if hvidx==1
```

```
gen n=1
```

```
* add up the number of households in each cluster  
collapse (sum) n, by(hv001)
```

```
summarize  
tab n
```

Here's the result. There were 140 clusters. The number of households per cluster ranged from 7 to 41, was 9 or 10 in a third of the clusters, and the mean was 15.2 households per cluster.

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
hv001	140	103.85	73.81436	1	248
n	140	15.15714	7.107875	7	41

```
.
```

```
. tab n
```

(sum) n	Freq.	Percent	Cum.
-----+-----			
7	2	1.43	1.43
8	4	2.86	4.29
9	22	15.71	20.00
10	25	17.86	37.86
11	9	6.43	44.29
12	5	3.57	47.86
13	8	5.71	53.57
14	6	4.29	57.86
15	9	6.43	64.29

16	8	5.71	70.00
17	3	2.14	72.14
18	4	2.86	75.00
19	3	2.14	77.14
20	3	2.14	79.29
21	3	2.14	81.43
22	5	3.57	85.00
23	4	2.86	87.86
25	4	2.86	90.71
26	1	0.71	91.43
27	1	0.71	92.14
28	2	1.43	93.57
30	2	1.43	95.00
31	1	0.71	95.71
33	1	0.71	96.43
34	1	0.71	97.14
35	3	2.14	99.29
41	1	0.71	100.00

---

Total | 140 100.00

If you replace the last two Stata lines with

```
collapse (sum) n, by(hv001 hv025)
tab hv025, summarize(n)
```

Then you get the following results separately by urban and rural areas.

type of   place of   residence	Summary of (sum) n		Freq.
	Mean	Std. Dev.	
urban	14.21	6.0607947	100
rural	17.525	8.875167	40
Total	15.157143	7.1078753	140