

---

Subject: What is the DHS position on the use of cluster-level data?

Posted by [ld190](#) on Sun, 27 Mar 2016 16:27:36 GMT

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

---

Dear All,

Apologies in advance for the long response and thank you for reading it. I've tried to be thorough in order to bypass old material and to get to what I see as the key ambiguities (for me) remaining in the question of how, and whether, cluster-level data from the DHS should be analyzed.

I have come a point in my research that am starting to use DHS data in earnest. I have downloaded and conducted some preliminary analysis of datasets from Senegal, and some of the things that I would like to do involve aggregating from individual to "cluster-level" characteristics. So for example, calculating that X percent of Women 15-49 sampled in cluster Y have attribute A and Z percent have B attribute. I am interested in clusters because my research is about residential communities - small areas where people are co resident and have some chance of knowing or being influenced by each other. I am able to do this using SPSS and aggregating cases by Cluster. This analysis produces some very interesting (and theoretically plausible) results. However, I am concerned about warnings that I've read, against the disaggregation of DHS data.

Despite attempting to work through the DHS's very helpful store of literature, and this forum, I remain unsure about the DHS' position on the direct use of cluster data. On the one hand, the official guides, and very enjoyable YouTube tutorials, seem to me to emphasise that the surveys are designed to be representative at the Regional and National levels only, meaning that further disaggregation is not possible. However, I'm not sure about the extent to which this applies to my research. I am not interested in estimating the prevalence of attribute A (which is very common) for any area except for the cluster (the Enumeration Area) itself. So I'm not interested in the surrounding administrative area or some other geographic area, for example. I'm just interested in the cluster of households from which the chosen households were randomly sampled. Considering this level of analysis, around 20 households sampled at random from a pool of, on average, 110 households, is the data so unrepresentative as to be useless? Do the observed attributes of the randomly sampled households (20) tell us nothing reliable about the attributes of the overall population (110)? What about if we average across a large number of clusters, to produce a distribution of values?

The guidance on this issue on the forum appears to me to provide a number of alternative possible answers and issues to consider.

One user on the forum seems to suggest that the use of cluster-level data is "noisy" (error prone) but basically OK as long as this is taken into account and that it is common to use this level of data for certain purposes:

[http://userforum.dhsprogram.com/index.php?t=msg&goto=9054&S=41b1f8e9c6ffff1e5ed1b91414054772&srch=aggregatin+g+clusters#msg\\_9054](http://userforum.dhsprogram.com/index.php?t=msg&goto=9054&S=41b1f8e9c6ffff1e5ed1b91414054772&srch=aggregatin+g+clusters#msg_9054)

However, DHS staff member Trevor, on another post suggests that the use of cluster-level estimates are "impossible" because the sample sizes are too small. Although he is referring to

calculating Child Mortality rates, which is a very rare event, and so this measure might require an especially large sample size.

[http://userforum.dhsprogram.com/index.php?t=msg&goto=8524&S=41b1f8e9c6fff1e5ed1b91414054772&srch=aggregatin+clusters#msg\\_8524](http://userforum.dhsprogram.com/index.php?t=msg&goto=8524&S=41b1f8e9c6fff1e5ed1b91414054772&srch=aggregatin+clusters#msg_8524)

On another post, Trevor says this:

Quote:You can and should still use hv005 as the sample weight, but doing your analysis with smaller geographic units is potentially problematic. The sample is designed to be representative at the region level, but not at the level of smaller units. As you disaggregate the data to smaller units the sample is less and less likely to be representative. The sample is also designed to provide a certain level of accuracy at the region level, and again as you disaggregate to smaller units the accuracy of those estimates gets worse and worse and the confidence intervals around the estimates quickly become very large and unreliable.

I found this advice slightly confusing. Presumably going from drawing inferences about the population at an officially representative level (region), to an intermediate level (like a small administrative unit) might reduce the representativeness of the data. This I because the size of the sample (N households) might be getting smaller relative to the size of target population (a whole administrative district). However, presumably at some point this trend will reverse? If we only tried to draw inferences about the Enumeration Area from which the sample is drawn, for example, then surely this is more representative than trying to use the cluster sample to draw inferences about, for example, a larger population within 5km<sup>2</sup> of the Enumeration Area?

ClaraB, also a DHS staff member, offers this advice on the interpretation of cluster-samples:

[http://userforum.dhsprogram.com/index.php?t=msg&goto=8315&S=41b1f8e9c6fff1e5ed1b91414054772&srch=aggregatin+clusters#msg\\_8315](http://userforum.dhsprogram.com/index.php?t=msg&goto=8315&S=41b1f8e9c6fff1e5ed1b91414054772&srch=aggregatin+clusters#msg_8315)

Quote:[inference about the] district location of the sampled clusters using a GIS software and the GPS dataset these data would not be statistically representative.

However, I'm unclear how to interpret this advice. Is the warning given because the user is trying to draw inferences about the district level (larger than the EA) from a single sample cluster?

Finally, a forum user posted this advice about the use of cluster-level measures:

Quote:cluster-level measurements are based on too few observations to be meaningful in and of themselves - as you say, there are wildly under-powered. A couple of things you could do: a) by averaging over many clusters, you can still get good estimates of community level variables, but each individual cluster-level point-estimate would be very, very noisy. But they may still mostly "agree" in some sense; b) so if in your hierarchical model you allow each cluster an unconstrained cluster-specific effect (like treating each cluster as a mini-experiment), you could look at those individual point-estimates on a scatter plot (say Beta across some variable you think would affect Beta); c) and then you could start restricting those Betas to have some particular distribution (a random slope model) and see how that changes your overall point estimate as you make your priors on the distribution of Beta more/less informative. I think this makes sense as a kind of

model-checking or informal/additional inference procedure. A leave-one-out cross-validation approach might make sense too, depending on how you end up thinking about each of these within-cluster estimates.

This user's scatter-plot suggestion is very close to what I have done in my own research.

In addition to searching the DHS forums, I've discovered that some published academic work has engaged with data at the cluster-level. Storey and Kaggwa from the Department of Population, Johns Hopkins University, have used cluster level data from the 1995, 2000 and 2005 Egypt Demographic and Health Surveys (EDHS).

This is a quote from the abstract for their paper:

Quote: Norms are defined at the cluster level, which serves as our community-level unit of analysis

The official site for the article is here:

[http://muse.jhu.edu/journals/population\\_review/v048/48.1.storey.html](http://muse.jhu.edu/journals/population_review/v048/48.1.storey.html)

Also there has also been some research to actually estimate the error introduced from using cluster-level measures with DHS data. This was conducted by Øystein Kravdal, Professor of Demography at the University of Oslo.

Here is a quote from the abstract for his paper:

Quote: For example, researchers may consider including in their models the average education within the sample (cluster) of approximately 25 women interviewed in each primary sampling unit (PSU). However, this is only a proxy for the theoretically more interesting average among all women in the PSU, and, in principle, the estimated effect of the sample mean may differ markedly from the effect of the latter variable. Fortunately, simulation experiments show that the bias actually is fairly small - less than 14% - when education effects on first birth timing are estimated from DHS surveys in sub-Saharan Africa. If other data are used, or if the focus is turned to other independent variables than education, the bias may, of course, be very different. In some situations, it may be even smaller; in others, it may be unacceptably large. That depends on the size of the clusters, and on how the independent variables are distributed within and across communities. Some general advice is provided.

This paper is available to read, published in a Peer Reviewed Open Source Journal:

<http://www.demographic-research.org/volumes/vol15/1/>

Both of these papers, seem favorable to the use of cluster-level DHS data.

I wonder if the 'proof of the pudding is in the eating'? The results from my analysis of community level data are theoretically plausible, there is a clear pattern (agreement) in a scatter plot showing the relationship between two measures (the frequency of observations A and B in each cluster) across all the clusters and this pattern is consistent across Senegalese DHS surveys in 2005, 2010 and 2014. Presumably, if the level of noise were so great that no meaningful information could be gained from cluster-level analysis, then a clear pattern of results like this would be quite

surprising?

Thank you again for reading through this long question. I am by no-means certain about any of this, I am new to this area of analysis and this kind of analysis. However I wanted to provide a detailed description of the problem that I am trying to grapple with.

If anyone can offer any further thoughts, clarification or advice on the use of cluster-level analysis with DHS data, I would be very grateful to hear it. Also, if there is some key DHS document (or other publication) that I have missed which elaborates on this issue would be grateful to receive a recommendation.

Many thanks in advance for your response.

Laurence.

P.s Thanks to UserRHS for the help in improving the formatting of this post.

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [user-rhs](#) on Sun, 27 Mar 2016 20:47:07 GMT

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

---

The key issue is that some of the clusters may be too small for the analysis to be meaningful. For example, how would you interpret the average from 3 households? You weren't clear in your original question about how you were using the aggregates, but since you cited the Storey & Kaggwa and Kravdal papers and quoted Reduced-For(u)m's responses, it looks like you want to enter them as covariates in a regression model.

There are no hard and fast rules about what to do, and determination should be done on a case-by-case basis and depends on: 1.) the overall sample size, 2.) the number of clusters, 3.) the number of observations in the cluster. If you have many clusters with a sufficient number of observations, your results will be less biased than if you many clusters with a small number of observations. For example, I would be fairly comfortable entering cluster-level averages into a model from the Indonesia 2012 DHS into a regression model, because 1.) it's huge (>45,000 observations), 2.) it has a lot of clusters (1,832 clusters), 3.) the clusters are sufficiently "large" (around 90% of the clusters have 20 or more people in them, and only about 80 people live in clusters with <10 observations each), but I would have less confidence doing it with a dataset with 5,000 observations and 1,200 clusters where the average size is 10 (I worked with a dataset like that once, and I ended up aggregating up to the district level to get respectable sizes) .

Second, cluster-level analysis can still be useful, depending on the level of inference. I think what the DHS team cautions against is making population-level inference based on the clusters, because the survey is not designed for that level of disaggregation. You can make a case for valid inference to the sample in the worst-case-scenario, or at least minimize the population-level implications of your findings.

Third, even the experts are still in disagreement about this, which works to your advantage. You can take one school of thought and justify it with citations from the peer-reviewed literature. At the end of the day, science is about weighing different opinions and evidence and defending your choices.

A good first step is determining the number and size of your clusters. If they are sufficiently "large" and you can make a case for it that's theoretically/empirically/clinically plausible, then why not? Cluster aggregates are less than ideal, but if we had better measures than cluster-level aggregates for whatever construct we were trying to operationalize, surely we would have used them instead of these proxies derived from the data, right? I would make the suggestion to fit the model first with just the individual/household-level variables first and enter the cluster-level aggregates separately to see how things change. Reduced-Form has some good advice, which you have quoted above.

NB: I'm not a DHS affiliate, so I can't offer the official DHS position

---

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [ld190](#) on Fri, 01 Apr 2016 10:47:13 GMT

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

---

Dear user-rhs,

First, many thanks for your detailed and very helpful response. I really appreciate the time-taken.

To address the ambiguity you noticed in my post: As it happens I'm actually not using regression models. I am using an agent-based model, which is a kind of computer simulation. In this case the cluster-level measures that I am interested in are for purposes of model calibration and validation (rather than "fitting" as in regression models). Cluster level measure X will be used to "calibrate" the simulation - which means it will be used to set the value of a key parameter in the simulation. Then measure Y will be used to validate (i.e. test) the simulation by observing whether the simulated relationship between input variable X, and the output of the simulation, are the same as the relationships which exists in the real data between measures X and Y. Hence, scatter plots showing the relationship between simulation-input X and the subsequent simulated-output (which should correspond to Y) can be overlayed with the real cluster-level X and Y values - as an indication of the match between the simulation and the real clusters. I am interested in cluster-level measures because the model is a flexible model of social dynamics within a cluster (a village-sized residential community with a social network - etc.).

Having said that, the advice about regression is very useful for future reference - thank-you.

Having thought about what you've written, and about the recommendations of the Kravdal paper, I am now cautiously optimistic. It seems that the analysis is worth pursuing for the moment. Kravdal

and you cite the importance of the absolute size of cluster samples (as well as the relative cluster population/ sample ratio). Having looked at the dataset from Senegal 2005, the cluster sizes are quite large ( $M = 38$ ,  $SD = 11$ ) and only 4.8% of clusters are below 20 cases. However, based on Kravdal's advice it will be important to check the within- and between- cluster variance of both measures. I'll also have to consider the average size of the cluster population itself. It may even be worth my creating a replication of something similar to Kravdal's simulation - in order to explore the viability of using this particular data-set in this way.

As and when I do a more in-depth investigation of the viability of using clusters for this purpose I will post about it here for the interest of future users of these particular measures.

Also thanks for the advice regarding a scientific justification for such choices. I agree that if a good justification can be found for the use of these measures and as long as one is open and honest about their limitations and drawbacks, there is no obligation to neglect their use.

Best,

Laurence.

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [user-rhs](#) on Mon, 04 Apr 2016 22:08:00 GMT

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

That's really neat, Laurence. I have never done agent-based modeling before, but I have former colleagues that do it for infectious disease (mainly STDs). Good luck in your studies. I'd be curious to see your results.

rhs

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [Lukresha](#) on Sun, 24 Jul 2016 14:34:46 GMT

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

I would like to calculate the percentage of women in each cluster that are in various occupations (not working, professional/technical/managerial, agriculture, household and domestic, skilled manual, unskilled manual)

Kindly, how can I go about clustering in stata as I tried using the collapse command but to no

avail.

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [Reduced-For\(u\)m](#) on Sun, 24 Jul 2016 22:51:45 GMT

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

---

Why didn't the "collapse" command work? Is it just because the "svy" prefix doesn't work with collapse? You may not need that, since all HH in the same cluster have the same weight, and since you aren't doing any inference calculations (p-values) you don't need to worry about the stratification either. If you just want cluster means, collapse should work just fine.

---

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [Lukresha](#) on Mon, 25 Jul 2016 09:33:44 GMT

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

---

used the command collapse mv717, by( hv001) where mv717 is occupation grouped and hv001 is cluster number.

This is the output I got:

(mean)   mv717	Freq.	Percent	Cum.
-----+-----			
0	40	4.29	4.29
.25	1	0.11	4.39
.2857143	1	0.11	4.50
.3	1	0.11	4.61
.3333333	2	0.21	4.82
.4444444	1	0.11	4.93
.5	3	0.32	5.25
.5714286	1	0.11	5.36
.6	1	0.11	5.47
.6666667	2	0.21	5.68
.7142857	1	0.11	5.79
.75	1	0.11	5.89
.8	2	0.21	6.11
.8571429	1	0.11	6.22
.8888889	1	0.11	6.32
.9230769	1	0.11	6.43
1	18	1.93	8.36
1.142857	4	0.43	8.79
1.222222	1	0.11	8.90
1.230769	1	0.11	9.00
1.25	2	0.21	9.22



1.272727	1	0.11	9.32
1.285714	1	0.11	9.43
1.307692	1	0.11	9.54
1.333333	7	0.75	10.29
1.428571	1	0.11	10.40
1.454545	1	0.11	10.50
1.5	3	0.32	10.83
1.555556	1	0.11	10.93
1.6	5	0.54	11.47
1.625	1	0.11	11.58
1.636364	1	0.11	11.68
1.684211	1	0.11	11.79
1.714286	1	0.11	11.90
1.75	1	0.11	12.00
1.777778	1	0.11	12.11
1.8	3	0.32	12.43
1.818182	2	0.21	12.65
1.833333	1	0.11	12.75
1.846154	2	0.21	12.97
1.857143	2	0.21	13.18
1.875	1	0.11	13.29
1.913043	1	0.11	13.40
1.928571	1	0.11	13.50
2	26	2.79	16.29
2.083333	1	0.11	16.40
2.086957	1	0.11	16.51
2.111111	1	0.11	16.61
2.117647	1	0.11	16.72
2.125	2	0.21	16.93
2.133333	1	0.11	17.04
2.142857	1	0.11	17.15
2.166667	2	0.21	17.36
2.181818	1	0.11	17.47
2.214286	1	0.11	17.58
2.222222	5	0.54	18.11
2.25	8	0.86	18.97
2.272727	2	0.21	19.19
2.285714	4	0.43	19.61
2.3	1	0.11	19.72
2.307692	1	0.11	19.83
2.333333	5	0.54	20.36
2.384615	1	0.11	20.47
2.4	6	0.64	21.11
2.409091	1	0.11	21.22
2.428571	4	0.43	21.65
2.444444	2	0.21	21.86
2.454545	1	0.11	21.97
2.466667	1	0.11	22.08



2.5	1	0.11	22.19
2.555556	1	0.11	22.29
2.571429	4	0.43	22.72
2.588235	1	0.11	22.83
2.6	2	0.21	23.04
2.615385	1	0.11	23.15
2.625	1	0.11	23.26
2.666667	13	1.39	24.65
2.714286	4	0.43	25.08
2.727273	1	0.11	25.19
2.75	1	0.11	25.29
2.777778	1	0.11	25.40
2.8	4	0.43	25.83
2.833333	1	0.11	25.94
2.857143	2	0.21	26.15
2.875	3	0.32	26.47
2.888889	2	0.21	26.69
2.9	1	0.11	26.80
2.909091	1	0.11	26.90
2.92	1	0.11	27.01
3	25	2.68	29.69
3.058824	1	0.11	29.80
3.076923	3	0.32	30.12
3.1	1	0.11	30.23
3.111111	1	0.11	30.33
3.125	1	0.11	30.44
3.142857	2	0.21	30.65
3.166667	2	0.21	30.87
3.2	4	0.43	31.30
3.222222	1	0.11	31.40
3.230769	3	0.32	31.73
3.25	2	0.21	31.94
3.272727	1	0.11	32.05
3.294118	1	0.11	32.15
3.3	1	0.11	32.26
3.315789	1	0.11	32.37
3.333333	8	0.86	33.23
3.375	2	0.21	33.44
3.4	4	0.43	33.87
3.428571	4	0.43	34.30
3.4375	2	0.21	34.51
3.444444	3	0.32	34.83
3.466667	1	0.11	34.94
3.473684	1	0.11	35.05
3.5	17	1.82	36.87
3.538461	1	0.11	36.98
3.545455	1	0.11	37.08
3.555556	5	0.54	37.62

3.571429	4	0.43	38.05
3.583333	2	0.21	38.26
3.6	3	0.32	38.59
3.611111	1	0.11	38.69
3.625	6	0.64	39.34
3.636364	1	0.11	39.44
3.666667	1	0.11	39.55
3.692308	2	0.21	39.76
3.7	1	0.11	39.87
3.705882	1	0.11	39.98
3.714286	1	0.11	40.09
3.722222	1	0.11	40.19
3.75	5	0.54	40.73
3.777778	3	0.32	41.05
3.785714	1	0.11	41.16
3.789474	1	0.11	41.26
3.8	2	0.21	41.48
3.833333	1	0.11	41.59
3.846154	1	0.11	41.69
3.857143	2	0.21	41.91
3.875	1	0.11	42.02
3.9	1	0.11	42.12
3.909091	2	0.21	42.34
3.928571	1	0.11	42.44
3.947368	1	0.11	42.55
3.954545	1	0.11	42.66
4	103	11.04	53.70
4.111111	1	0.11	53.80
4.125	2	0.21	54.02
4.133333	1	0.11	54.13
4.153846	1	0.11	54.23
4.166667	3	0.32	54.56
4.181818	2	0.21	54.77
4.2	3	0.32	55.09
4.25	1	0.11	55.20
4.266667	1	0.11	55.31
4.277778	1	0.11	55.41
4.285714	5	0.54	55.95
4.307693	1	0.11	56.06
4.333333	7	0.75	56.81
4.357143	1	0.11	56.91
4.363636	3	0.32	57.23
4.375	1	0.11	57.34
4.384615	1	0.11	57.45
4.388889	1	0.11	57.56
4.4	3	0.32	57.88
4.428571	2	0.21	58.09
4.444445	1	0.11	58.20

4.454545	2	0.21	58.41
4.5	11	1.18	59.59
4.538462	2	0.21	59.81
4.555555	1	0.11	59.91
4.5625	1	0.11	60.02
4.571429	3	0.32	60.34
4.615385	2	0.21	60.56
4.625	1	0.11	60.66
4.666667	6	0.64	61.31
4.692307	1	0.11	61.41
4.7	1	0.11	61.52
4.714286	1	0.11	61.63
4.727273	1	0.11	61.74
4.75	6	0.64	62.38
4.769231	1	0.11	62.49
4.777778	1	0.11	62.59
4.8	11	1.18	63.77
4.8125	1	0.11	63.88
4.818182	1	0.11	63.99
4.833333	2	0.21	64.20
4.846154	1	0.11	64.31
4.857143	1	0.11	64.42
4.875	1	0.11	64.52
4.909091	2	0.21	64.74
4.9375	2	0.21	64.95
5	22	2.36	67.31
5.052631	1	0.11	67.42
5.055555	1	0.11	67.52
5.0625	1	0.11	67.63
5.076923	1	0.11	67.74
5.090909	4	0.43	68.17
5.1	1	0.11	68.27
5.142857	5	0.54	68.81
5.166667	3	0.32	69.13
5.181818	1	0.11	69.24
5.2	8	0.86	70.10
5.235294	1	0.11	70.20
5.25	3	0.32	70.53
5.333333	8	0.86	71.38
5.384615	2	0.21	71.60
5.4	3	0.32	71.92
5.428571	2	0.21	72.13
5.444445	1	0.11	72.24
5.454545	1	0.11	72.35
5.5	13	1.39	73.74
5.545455	1	0.11	73.85
5.555555	2	0.21	74.06
5.6	3	0.32	74.38

5.625	2	0.21	74.60
5.666667	10	1.07	75.67
5.7	1	0.11	75.78
5.714286	3	0.32	76.10
5.727273	1	0.11	76.21
5.733333	1	0.11	76.31
5.75	1	0.11	76.42
5.785714	1	0.11	76.53
5.8	1	0.11	76.63
5.833333	1	0.11	76.74
5.857143	1	0.11	76.85
5.888889	2	0.21	77.06
5.909091	1	0.11	77.17
6	33	3.54	80.71
6.111111	3	0.32	81.03
6.125	1	0.11	81.14
6.142857	1	0.11	81.24
6.153846	1	0.11	81.35
6.166667	1	0.11	81.46
6.2	3	0.32	81.78
6.222222	1	0.11	81.89
6.3	1	0.11	81.99
6.307693	1	0.11	82.10
6.333333	5	0.54	82.64
6.363636	2	0.21	82.85
6.375	1	0.11	82.96
6.4	2	0.21	83.17
6.421052	1	0.11	83.28
6.428571	3	0.32	83.60
6.5	14	1.50	85.10
6.545455	2	0.21	85.32
6.583333	1	0.11	85.42
6.6	2	0.21	85.64
6.642857	1	0.11	85.74
6.666667	1	0.11	85.85
6.692307	2	0.21	86.07
6.714286	4	0.43	86.50
6.75	3	0.32	86.82
6.8	2	0.21	87.03
6.857143	2	0.21	87.25
6.888889	2	0.21	87.46
6.909091	1	0.11	87.57
7	10	1.07	88.64
7.076923	1	0.11	88.75
7.090909	2	0.21	88.96
7.111111	4	0.43	89.39
7.142857	1	0.11	89.50
7.181818	1	0.11	89.60

7.2	2	0.21	89.82
7.214286	1	0.11	89.92
7.230769	1	0.11	90.03
7.25	1	0.11	90.14
7.307693	1	0.11	90.25
7.333333	8	0.86	91.10
7.4	1	0.11	91.21
7.416667	1	0.11	91.32
7.5	3	0.32	91.64
7.538462	1	0.11	91.75
7.555555	1	0.11	91.85
7.571429	1	0.11	91.96
7.583333	1	0.11	92.07
7.636364	1	0.11	92.18
7.666667	1	0.11	92.28
7.714286	3	0.32	92.60
7.75	1	0.11	92.71
7.857143	2	0.21	92.93
7.888889	1	0.11	93.03
8	16	1.71	94.75
8.142858	1	0.11	94.86
8.166667	3	0.32	95.18
8.25	2	0.21	95.39
8.285714	1	0.11	95.50
8.4	2	0.21	95.71
8.6	2	0.21	95.93
8.608696	1	0.11	96.03
8.785714	1	0.11	96.14
8.888889	1	0.11	96.25
9	35	3.75	100.00
-----+-----			
Total	933	100.00	

From this output I cannot tell apart the occupation categories. Before collapse, these were my occupation categories:

tab mv717

occupation (grouped)	Freq.	Percent	Cum.
-----+-----			
not working	1,865	24.72	24.72
professional/technical/managerial	515	6.83	31.55
agriculture - self employed	2,507	33.23	64.78
household and domestic	978	12.96	77.74
skilled manual	527	6.99	84.73
unskilled manual	1,152	15.27	100.00
-----+-----			
Total	7,544	100.00	

---

Subject: Re: What is the DHS position on the use of cluster-level data?

Posted by [Reduced-For\(u\)m](#) on Mon, 25 Jul 2016 22:00:51 GMT

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

---

See response here: [http://userforum.dhsprogram.com/index.php?t=tree&th=5328&goto=10400&#msg\\_10400](http://userforum.dhsprogram.com/index.php?t=tree&th=5328&goto=10400&#msg_10400)

---