Subject: SSU variables in the Mauritania DHS2019-2021
Posted by kinden on Thu, 29 Feb 2024 18:56:26 GMT
View Forum Message <> Reply to Message

Dear DHS team,

According to the DHS Household and Sampling Manual, the DHS surveys are two stage surveys. The sampling units for the first stage of selection is called the Primary Sampling Unit (PSU) with probability sampling of EAs; the sampling unit for the second stage of selection is called the Secondary Sampling Unit (SSU) with random sampling of households in the selected EAs.

I try to extract the data to identify the following categories of pregnant women who delivered their last child in the past five years preceding to the Mauritania DHS2019-2021 survey.
(1) The number of pregnant women who delivered their last child after August 2019 in the regions of Hodh Gharbi and Guidimakha
(2) The number of pregnant women who delivered their last child before August 2019 in the regions of Hodh Gharbi and Guidimakha
(3) The number of pregnant women who delivered their last child after August 2019 in other regions than Hodh Gharbi and Guidimakha
(4) The number of pregnant women who delivered their last child before August 2019 in other regions than Hodh Gharbi and Guidimakha

However, I could only get the data regarding the number of women in the sample in the PSU: 39,793 women in 39,793 households.
What is the secondary sampling unit(ssu) variable in DHS data set?
My objective is not to produce sampling error but rather to identify the sample size of the above-mentioned different categories of pregnant women.

Any suggestions would be appreciated.
Thank you!

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by Bridgette-DHS on Fri, 01 Mar 2024 12:45:26 GMT
View Forum Message <> Reply to Message

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

The variable to identify PSUs is v001 (duplicated as v021) and the variable for households, which are the SSUs, is v002. However, you don't need either of them. The following lines will give a table for the number of births that the sampled women had before August 2019, during August 2019, or after August 2019, for each region (v024), not just the two regions you mentioned.

Note that these are numbers of births to women in the sample. The number of births after August 2019 is affected by when the woman was interviewed. The fieldwork started in November 2019 and was stopped in March 2020 (there were 2 interviews in April 2020), and then resumed for two months in February-March 2021. If you are trying to get at the impact of COVID-19 on fertility, you will have to take the date of interview into account. Only the 5,016 women interviewed in 2021

could possibly show a COVID-19 effect.

* Number of births in the Mauritania 2019-21 survey before/after August 2019, by region

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\MRBR71FL.DTA", clear

* Find the cmc for August 2019
summarize b3 if b1==8 & b2==2019
* August 2019 is b3=1436

* Construct a variable for the timing of births in the birth histories
gen birthdate=1
replace birthdate=2 if b3==1436
replace birthdate=3 if b3>1436

label define bd 1 "Before August 2019" 2 "In August 2019" 3 "After August 2019"
label values birthdate bd
tab v024 birthdate [iweight=v005/1000000]

Here is the table of weighted frequencies:

birthdate		
region	Before Au	In August	After Aug	Total
hodh echargui	5,398.926	40.237882	165.757235	5,604.9211
hodh gharbi	4,306.5666	28.994562	137.206754	4,472.768
assaba	3,542.996	19.488518	85.435286	3,647.92
gorgol	3,866.476	21.88072	100.96081	3,989.317
brakna	2,995.976	21.032225	97.846587	3,114.855
trarza	1,992.851	16.385234	51.722557	2,060.959
adrar	676.115868	2.597507	16.480042	695.193417
dakhlet nouadhibou	1,097.09	4.328784	28.115727	1,129.534
tagant	867.836632	5.258122	23.18794	896.282694
guidimagha	3,860.71	23.182374	104.898378	3,988.791
tiris zemour et inchi	589.395148	6.91581	15.450279	611.761237
nouakchott ouest	1,339.084	8.220457	146.107834	1,493.412
nouakchott nord	4,009.97	17.350836	347.069054	4,374.39
nouakchott sud	3,567.589	18.178068	302.901025	3,888.668
Total	38,111.58	234.051099	1,623.14	39,968.77

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by kinden on Thu, 07 Mar 2024 17:20:44 GMT

Thank you so much, Bridgette.
Your response is very helpful.

Yet, the table below shows the number of livebirths.
How can I get the number of the last child born before/after August 2019 among the classified pregnant women by region?

Could you also confirm the collected data (5,016 women interviewed) in 2021 were only from Nouakchott as the DHS report says?

Kind regards,
Kazumi

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by Bridgette-DHS on Mon, 18 Mar 2024 12:15:03 GMT

Following is a response from Senior DHS staff member, Tom Pullum:

We apologize for the delay in this response.

The variable for region, v024, has the following categories:

. label list V024
V024:
1 hodh echargui
2 hodh gharbi
3 assaba
4 gorgol
5 brakna
6 trarza
7 adrar
8 dakhlet nouadhibou
9 tagant
10 guidimagha
11 tiris zemour et inchiri
12 nouakchott ouest
13 nouakchott nord
14 nouakchott sud

Nouakshott is divided into west, north, and south. The weighted and unweighted numbers of women in the regions is obtained with the IR file:

. use "...MRIR71FL.DTA"

. tab v024

region	Freq.	Percent	Cum.
hodh echargui	1,327	8.44	8.44
hodh gharbi	1,402	8.92	17.37
assaba	1,380	8.78	26.15
gorgol	1,449	9.22	35.37
brakna	1,454	9.25	44.62
trarza	1,320	8.40	53.02
adrar	808	5.14	58.16
dakhlet nouadhibou	643	4.09	62.26
tagant	920	5.85	68.11
guidimagha	1,786	11.37	79.48
tiris zemour et inchiri	787	5.01	84.49
nouakchott ouest	728	4.63	89.12
nouakchott nord	910	5.79	94.91
nouakchott sud	800	5.09	100.00
Total	15,714	100.00	

. tab v024 [iweight=v005/1000000]

region	Freq.	Percent	Cum.
hodh echargui	2,033.68582	12.94	12.94
hodh gharbi	1,579.5074	10.05	22.99
assaba	1,249.2951	7.95	30.94
gorgol	1,292.7495	8.23	39.17
brakna	1,281.9262	8.16	47.33
trarza	961.483433	6.12	53.45
adrar	297.682976	1.89	55.34
dakhlet nouadhibou	538.786626	3.43	58.77
tagant	349.098656	2.22	60.99
guidimagha	1,243.5337	7.91	68.91
tiris zemour et inchiri	273.47023	1.74	70.65
nouakchott ouest	792.569719	5.04	75.69
nouakchott nord	2,073.0319	13.19	88.88
nouakchott sud	1,747.1787	11.12	100.00
Total	15,714	100.00	

However, in order to get the 5016 that you refer to, I have to go to the BR file, in which the cases

are all the children in the birth histories:

```
. use "...MRBR71FL.DTA", clear

. tab v024

                region |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
         hodh echargui |      3,560        8.95        8.95
           hodh gharbi |      3,865        9.71       18.66
                assaba |      3,902        9.81       28.46
                gorgol |      4,355       10.94       39.41
                brakna |      3,467        8.71       48.12
                trarza |      2,841        7.14       55.26
                 adrar |      1,832        4.60       59.86
    dakhlet nouadhibou |      1,388        3.49       63.35
                tagant |      2,300        5.78       69.13
            guidimagha |      5,467       13.74       82.87
tiris zemour et inchiri |      1,800        4.52       87.39
       nouakchott ouest |      1,279        3.21       90.61
       nouakchott nord |      1,936        4.87       95.47
        nouakchott sud |      1,801        4.53      100.00
-----------------------+-----------------------------------
                 Total |     39,793      100.00

. tab v024 [iweight=v005/1000000]

                region |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
         hodh echargui | 5,604.9211       14.02       14.02
           hodh gharbi | 4,472.7679       11.19       25.21
                assaba | 3,647.9203        9.13       34.34
                gorgol | 3,989.3172        9.98       44.32
                brakna | 3,114.8548        7.79       52.12
                trarza | 2,060.9592        5.16       57.27
                 adrar | 695.193417        1.74       59.01
    dakhlet nouadhibou | 1,129.5344        2.83       61.84
                tagant | 896.282694        2.24       64.08
            guidimagha |3,988.79101        9.98       74.06
tiris zemour et inchiri | 611.761237        1.53       75.59
       nouakchott ouest |  1,493.412        3.74       79.33
       nouakchott nord |4,374.38955       10.94       90.27
        nouakchott sud | 3,888.6678        9.73      100.00
-----------------------+-----------------------------------
                 Total | 39,968.772      100.00
```

By adding the numbers in regions 12, 13, and 14, weighted and unweighted, in the IR and BR files, I found that the 5016 is the sum of the weighted frequencies in the BR file. I have highlighted those numbers in yellow. In other words, I can confirm that there are 5016 BIRTHS, NOT WOMEN, and these are WEIGHTED frequencies, NOT UNWEIGHTED.

You also ask for "the number of the last child born before/after August 2019 among the classified pregnant women by region". I have tried to understand your question but can you clarify it? Are you trying to match a table in the final report, or are you doing something new?

As before, I do not know why you are particularly interested in August 2019, and I don't know whether you intend to include that month 2019 in "before" or "after". I don't know whey you are limiting yourself to women who are pregnant, which I interpret to mean pregnant at the time of the survey.

If I interpret your question literally, you want to restrict yourself to women who are pregnant at the time of the survey (v213=1), and to the most recent birth (bidx=1) and you want to know how many of those births were before or after August 2019, by region. Here are the Stata lines for doing that. Let us know if you want something else.

```
use "...MRBR71FL.DTA", clear

* Find the cmc for August 2019
summarize b3 if b1==8 & b2==2019
* August 2019 is b3=1436

* Construct a variable for the timing of births in the birth histories
gen birthdate=1
replace birthdate=2 if b3==1436
replace birthdate=3 if b3>1436

label define bd 1 "Before August 2019" 2 "In August 2019" 3 "After August 2019"
label values birthdate bd

tab v024 birthdate if bidx==1 & v213==1 [iweight=v005/1000000]
```

```
. tab v024 birthdate if bidx==1 & v213==1 [iweight=v005/1000000]

                |           birthdate
         region | Before Au  In August  After Aug |     Total
----------------+---------------------------------+----------
   hodh echargui |149.223643   2.148514          0 |151.372157
    hodh gharbi |151.271238          0          0 |151.271238
          assaba | 76.345168          0          0 | 76.345168
          gorgol |102.743707          0          0 |102.743707
          brakna | 42.965639          0          0 | 42.965639
          trarza | 53.324355          0          0 | 53.324355
           adrar | 13.528143          0          0 | 13.528143
```

```
   dakhlet nouadhibou | 26.032095          0          0 | 26.032095
              tagant | 23.939958          0          0 | 23.939958
           guidimagha | 81.070053          0    .723028 | 81.793081
tiris zemour et inchi | 15.290715          0          0 | 15.290715
       nouakchott ouest | 17.218033          0    .792849 | 18.010882
        nouakchott nord | 91.418925   2.031406   6.741004 |100.191335
        nouakchott sud | 83.416651    5.80855   5.440926 | 94.666127
----------------------+---------------------------------+----------
                Total |927.788323    9.98847  13.697807 |  951.4746
```

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by kinden on Wed, 27 Mar 2024 21:45:54 GMT
View Forum Message <> Reply to Message

Thank you so much once again, Bridgette.
It is very useful to confirm that 5016 is the sum of the weighted frequencies in the BR file.

And yes, I would like to restrict the data to women who are pregnant at the time of the survey (v213=1), and to the most recent birth (bidx=1) and I want to know how many of those births were before or after August 2019, by region, assuming that most pregnant women delivered one baby at their last birth.
The reasons why I am interested in before/after August 2019 are that the intervention of my interest has started since December 2018 (9 months before August 2019) and that I would like to analyze key issues around ANC service delivery.

What I still do not understand from your two useful responses is why the number of livebirths after August 2019 is different between the data from the first response (1,623.14 - weighted) and those from the second response (13.697 - unweighted?)?
These numbers are supposed to be similar as no women will be able to deliver another baby from August 2019 onwards during the survey period, non?

* Number of births in the Mauritania 2019-21 survey before/after August 2019, by region

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\MRBR71FL.DTA", clear

* Find the cmc for August 2019
summarize b3 if b1==8 & b2==2019
* August 2019 is b3=1436

* Construct a variable for the timing of births in the birth histories
gen birthdate=1
replace birthdate=2 if b3==1436
replace birthdate=3 if b3>1436
```

```
label define bd 1 "Before August 2019" 2 "In August 2019" 3 "After August 2019"
label values birthdate bd
tab v024 birthdate [iweight=v005/1000000]
```

Here is the table of weighted frequencies:

```
                |           birthdate
         region | Before Au  In August  After Aug |     Total
----------------+---------------------------------+---- -------
  hodh echargui | 5,398.926  40.237882 165.757235 |5,604.9211
   hodh gharbi |4,306.5666  28.994562 137.206754 | 4,472.768
         assaba | 3,542.996  19.488518  85.435286 |  3,647.92
         gorgol | 3,866.476   21.88072  100.96081 | 3,989.317
         brakna | 2,995.976  21.032225  97.846587 | 3,114.855
         trarza | 1,992.851  16.385234  51.722557 | 2,060.959
          adrar |676.115868   2.597507  16.480042 |695.193417
dakhlet nouadhibou |  1,097.09   4.328784  28.115727 | 1,129.534
         tagant |867.836632   5.258122   23.18794 |896.282694
      guidimagha |  3,860.71  23.182374 104.898378 | 3,988.791
tiris zemour et inchi |589.395148    6.91581  15.450279 |611.761237
  nouakchott ouest | 1,339.084   8.220457 146.107834 | 1,493.412
  nouakchott nord |  4,009.97  17.350836 347.069054 |  4,374.39
   nouakchott sud | 3,567.589  18.178068 302.901025 | 3,888.668
----------------+---------------------------------+---- -------
          Total | 38,111.58 234.051099    1,623.14| 39,968.77
```

```
use "...MRBR71FL.DTA", clear

* Find the cmc for August 2019
summarize b3 if b1==8 & b2==2019
* August 2019 is b3=1436

* Construct a variable for the timing of births in the birth histories
gen birthdate=1
replace birthdate=2 if b3==1436
replace birthdate=3 if b3>1436

label define bd 1 "Before August 2019" 2 "In August 2019" 3 "After August 2019"
label values birthdate bd

tab v024 birthdate if bidx==1 & v213==1 [iweight=v005/1000000]
```

```
. tab v024 birthdate if bidx==1 & v213==1 [iweight=v005/1000000]

                |           birthdate
```

region	Before Au	In August	After Aug	Total
hodh echargui	149.223643	2.148514	0	151.372157
hodh gharbi	151.271238	0	0	151.271238
assaba	76.345168	0	0	76.345168
gorgol	102.743707	0	0	102.743707
brakna	42.965639	0	0	42.965639
trarza	53.324355	0	0	53.324355
adrar	13.528143	0	0	13.528143
dakhlet nouadhibou	26.032095	0	0	26.032095
tagant	23.939958	0	0	23.939958
guidimagha	81.070053	0	.723028	81.793081
tiris zemour et inchi	15.290715	0	0	15.290715
nouakchott ouest	17.218033	0	.792849	18.010882
nouakchott nord	91.418925	2.031406	6.741004	100.191335
nouakchott sud	83.416651	5.80855	5.440926	94.666127
Total	927.788323	9.98847	13.697807	951.4746

Kind regards,
Kazumi

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by Bridgette-DHS on Thu, 28 Mar 2024 17:07:34 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

The table with 1, 623.14 births after August 2019 was constructed with this command: "tab v024 birthdate [iweight=v005/1000000]". The table with just 13.697 births after August 2019 was constructed with "tab v024 birthdate if bidx==1 & v213==1 [iweight=v005/1000000]". Both frequencies were weighted, but the second and much smaller frequency was limited to the most recent birth (bidx=1) and to women who are pregnant at the time of the survey (v213=1). They are a very small subset of the births in the first table. These tables were what I understood you to be requesting.

You say that you are trying to measure the potential impact of an intervention related to ANC care. I don't see how either of these tables could be used to describe such an impact. This is an interesting question, but this survey may not be appropriate for answering it. I would like to help but the question is outside the scope of the user forum. I hope other users can help.

Here is a journal article that may help, but it uses two successive surveys:

Mallick, Lindsay, Trinadh Dontamsetti, Thomas Pullum, and Julia Fleuret. 2019. Using the Uganda Demographic and Health Surveys from 2011 and 2016 to assess changes in Saving Mothers, Giving Life intervention districts. J. of Global Health Research 3. doi:10.29392/joghr.3.e2019026.

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by kinden on Fri, 29 Mar 2024 17:08:18 GMT
View Forum Message <> Reply to Message

Thank you so much for your prompt reply.
I do understand that this is not the forum to conduct an additional analysis to measure the potential impact of an intervention related to ANC care.

What I am trying to do here is to verify the feasibility of my planned analysis on ANC service delivery.
If there are zero livebirth after August 2019 in most regions except Nouakchott, my analysis will not be feasible.
Thus, I have asked to clarify the differences of the previous two responses.

Kind regards,
Kazumi

Subject: Re: SSU variables in the Mauritania DHS2019-2021
Posted by Bridgette-DHS on Mon, 01 Apr 2024 12:20:45 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

My previous response said what was the difference between the two tables. I'm still not quite sure what your hypothesis is, about the difference between Nouakchott and the rest of Mauritania, but the comparisons that I believe you are trying to make would have to take account of the date of interview, which affects the length of time after August 2019 that the woman could have had a birth.

As you would know, the data collection for Mauritania was spread over 3 calendar years, 2019-2021. Most of the fieldwork was between late November 2019 and March of 2020. Because of Covid, fieldwork was suspended, and was not resumed and completed until February-March 2021. As it happens, the three areas that make up Nouakchott were the only areas that were surveyed in 2021.

Here is a table that gives the century month code (cmc, hv008) and the number of interviews in each month of data collection. This table has households as units and is unweighted. Ignore the

"totals" row and column:

. tab hv006 hv007, summarize(hv008) means freq

Means and Frequencies of date of interview (cmc)

month of interview	year of interview			
	2019	2020	2021	Total
1	.	1441	.	1441
	0	2540	0	2540
2	.	1442	1454	1443.591
	0	2545	389	2934
3	.	1443	1455	1447.8632
	0	2009	1369	3378
11	1439	.	.	1439
	115	0	0	115
12	1440	.	.	1440
	2691	0	0	2691
Total	1439.959	1441.9251	1454.7787	1443.3902
	2806	7094	1758	11658

The cmc's for February-March 2021 are 1454 and 1455. In those months, 389 and 1369 households, respectively, were interviewed.

Next, still with households as units, I get the number of interviews by region and cmc as follows:

tab hv024 hv008

	date of interview (cmc)							
region	1439	1440	1441	1442	1443	1454	1455	Total
hodh echargui	14	498	281	218	136	0	0	1,147
hodh gharbi	6	133	266	402	219	0	0	1,026
assaba	9	325	279	283	198	0	0	1,094
gorgol	14	263	263	267	129	0	0	936
brakna	6	246	298	263	239	0	0	1,052
trarza	15	254	275	292	190	0	0	1,026
adrar	7	230	157	134	140	0	0	668
dakhlet nouadhibou	8	141	147	141	231	0	0	668
tagant	14	226	241	197	51	0	0	729
guidimagha	16	279	169	118	245	0	0	827
tiris zemour et inchi	6	96	164	230	231	0	0	727

nouakchott ouest	0	0	0	0	0	103	479	582
nouakchott nord	0	0	0	0	0	174	417	591
nouakchott sud	0	0	0	0	0	112	473	585
Total	115	2,691	2,540	2,545	2,009	389	1,369	11,658

This table shows that, as I said, Nouakchott was not visited until 2021. For that reason alone, you would expect more births there in the interval since August 2019 than in the parts of Mauritania that were visited in late 2019 and early 2020.

To me, it is notable that the date collection is neatly divided, geographically, into pre-covid and post-covid, giving a kind of natural experimental design for potential effects of covid. But I don't see this as a good design for assessing the impact of an intervention in August 2019 in Nouakchott. Any potential effect of that intervention will be completely confounded with the potential effect of covid during the interruption to fieldwork.

If you have other questions, perhaps other users can help.