**Household Sampling – In-class example for RStudio data analysis of the 2009 Kenya DHS**

For this exercise, we will use the following dataset, which can be found under the “R Studio Lab” folder under Files. It can also be accessed directly in the GitHub repository titled “IHSD\_7440\_HH\_Sampling”

* 2009 Kenya DHS household-level dataset, called ***Kenya2009\_household\_1.csv***

We will now conduct some descriptive analyses on this dataset, given a few different assumptions as to how our data was sampled/collected. All relevant syntax, commands, and descriptions are featured in the R Markdown file provided in the “IHSD 7440 HH Sampling GitHub repository”. Please reference this for a step-by-step guide as to how to complete this in-class exercise.

***To start off with, we will as a few preliminary questions:***

1. What is your element in this analysis and how many are there (n)?

9057

1. How many clusters are there in this sample?

16

1. How many survey domains are there in this dataset?

8

Now complete the following tables. The tables show the proportion of households with electricity. For column d, you need to analyze the data appropriately, taking into account the following: 1) the use of a 2-stage cluster design that results in correlated data at the cluster / PSU level; 2) adjustment for differences in the ultimate probability of selection through sampling weights; and 3) uses the strata information to improve the precision of your estimates.

**In each cell, include the proportion and the standard error (round to 2 decimal points)**

Table 1: Proportion of households with electricity (variable HV206, which is labeled ‘Has electricity’)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = 9,057 | (a) Assuming SRS | (b) Assuming SRS, with weights | (c) 2-stage cluster sampling with weights | (d) 2-stage cluster sampling with weights and stratification |
| Residence () |  |  |  |  |
| Urban | 62.89 (0.90) | 65.64 (1.64) | 65.64 (4.05) | 65.64 (3.88) |
| Rural | 7.01 (0.33) | 8.11 (0.50) | 8.11 (1.44) | 8.11 (1.43) |
| Education Level () |  |  |  |  |
| None | 6.86(0.56) | 7.03(1.16) | 7.03(1.28) | 7.03(1.20) |
| Primary | 12.44(0.53) | 10.77(0.77) | 10.77(1.41) | 10.77(1.29) |
| Secondary | 39.87(1.08) | 38.69(1.78) | 38.69(4.24) | 34.69(3.97) |
| Higher | 77.99(1.29) | 72.68(2.03) | 72.68(3.54) | 72.68(3.05) |
| **All Households** | **24.96 (0.45)** | **23.04 (0.72)** | **23.04 (2.49)** | **23.04 (2.17)** |

**A few questions on this analysis-**

* + - What is the effect of sample weights on point estimates and standard errors?
    - What is the effect of the cluster sampling design (i.e. use of clusters at first stage) on point estimates and standard errors?
    - What is the design effect for this 2-stage cluster sampling design for the proportion of households with electricity, with sample weights and stratification included in the analysis?
    - How does household residence – urban versus rural - affect the proportion of households with at electricity?
    - Which of the four estimates (a, b, c, d) provides the least biased point estimate and standard error of for the proportion of households with electricity in Kenya?

***Now complete the following tables on your own.***

Table 2: Proportion of households that treat their water (variable HV237, which is labeled ‘Anything done to water to make safe to d’)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = 9,057 | (a) Assuming SRS | (b) Assuming SRS, with weights | (c) 2-stage cluster sampling with weights | (d) 2-stage cluster sampling with weights and stratification |
| Residence () |  |  |  |  |
| Urban |  |  |  |  |
| Rural |  |  |  |  |
| Education () |  |  |  |  |
| None |  |  |  |  |
| Primary |  |  |  |  |
| Secondary |  |  |  |  |
| Higher |  |  |  |  |
| **All Households** |  |  |  |  |

Table 3: Of those households that treat their water (variable HV237, which is labeled ‘Anything done to water to make safe to d’), what proportion tuse chlorine or bleach (variable hv237B, labeled ‘Water usually treated by: add bleach/chl’)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = | (a) Assuming SRS | (b) Assuming SRS, with weights | (c) 2-stage cluster sampling with weights | (d) 2-stage cluster sampling with weights and stratification |
| **All Households** |  |  |  |  |

Table 4: Proportion of households that had their house sprayed with insecticide (IRS) (variable SH125A, labeled ‘Sprayed interior walls for mosquitoes’)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = 9,057 | (a) Assuming SRS | (b) Assuming SRS, with weights | (c) 2-stage cluster sampling with weights | (d) 2-stage cluster sampling with weights and stratification |
| Residence () |  |  |  |  |
| Urban |  |  |  |  |
| Rural |  |  |  |  |
| Education () |  |  |  |  |
| None |  |  |  |  |
| Primary |  |  |  |  |
| Secondary |  |  |  |  |
| Higher |  |  |  |  |
| **All Households** |  |  |  |  |

Table 5: Mean number of mosquito nets per households (variable HML1, labeled ‘Number of mosquito nets’)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = 9,057 | (a) Assuming SRS | (b) Assuming SRS, with weights | (c) 2-stage cluster sampling with weights | (d) 2-stage cluster sampling with weights and stratification |
| Residence () |  |  |  |  |
| Urban |  |  |  |  |
| Rural |  |  |  |  |
| Education () |  |  |  |  |
| None |  |  |  |  |
| Primary |  |  |  |  |
| Secondary |  |  |  |  |
| Higher |  |  |  |  |
| **All Households** |  |  |  |  |