MICS6 - 10 - WS.1.1

Reid Minto

March 25, 2020

# Load libraries

library(tidyverse); library(haven); library(lubridate);  
library(knitr); library(kableExtra)  
library(Hmisc); library(survey)  
library(expss); library(here); library(gmodels)  
  
library(tableone); library(sjPlot); library(srvyr)

# Import data

here function navigates you to top level Rproj directory

# haven library called above has read\_sav function for reading in .sav / .sps etc  
# files from other languages  
here()

## [1] "C:/Users/Owner/Documents/UNICEF-MICS"

hh <- read\_sav(here("Data/Tunisia\_hh.sav"))  
# hh <- as\_factor(hh, only\_labelled = TRUE)

Filter only cases where hh interview = 1, create a variable hhweightHH48 by multiplying HH48 and hhweight, number of hh members by sample weight

Note: this is just first pass rough draft / run through - much of this looks like it can be consolidated into a single chunk / a single function execute per section from what I’ve seen so far, but broken up in first pass

hh <- hh %>% filter(HH46 == 1) %>%  
 mutate(hhweightHH48 = HH48 \* hhweight,  
 nhhmem = 1)

Include MICS6 - 10 - WS.sps is called

hh <- hh %>%   
 mutate(drinkingWater = ifelse(WS1 %in% c(11:14, 21, 31, 41, 51,   
 61, 71, 72, 91, 92), 1, 2))  
  
# `casewhen` cleaner than nested chains of ifelse for >2 conditions  
# same as SQL 'case when', but TRUE ~ val acts in place of "else"  
hh <- hh %>%  
 mutate(toiletType = case\_when(  
 WS11 %in% c(11:13, 18, 21, 22, 31) ~ 1,  
 WS11 == 95 ~ 3,  
 TRUE ~ 2  
 ))  
   
  
  
hh <- hh %>% mutate(flush = NA) %>%   
 mutate(flush = ifelse(WS11 %in% c(11:14, 18), 1, flush))  
  
  
  
hh = hh %>% mutate(sharedToilet = case\_when(  
 WS17 %in% c(1:5) ~ 1,  
 WS17 %in% 97:99 ~ 9,  
 is.na(WS17) ~ 0,  
 TRUE ~ 2))  
hh = hh %>% mutate(sharedToilet = ifelse(WS16 == 2, 3, sharedToilet))  
  
  
# visual check  
aa = hh %>% select(WS1, HH1, HH2, nhhmem, hhweightHH48, hhweight, HH48, stratum, PSU, drinkingWater, WS11)

After 10-WS include file call, resume

hh <- hh %>%   
 mutate(improvedWater = ifelse(drinkingWater == 1, 100, 0),  
 total = 1,  
 total100 = 100)

# All Value labels

Variable and Value labels all at once

Question if this is necessary at all in R, b/c can just as easily label R tables / figured with a list of labels when generating the tables / figures / reports, avoiding doing as we go along.

hh <- hh %>% apply\_labels(  
 nhhmem = "Number of household members",  
 WS1 = c(  
 "Tanker Truck" = 61,   
 "Cart with small tank" = 71,   
 "Bottled water [A]" = 91,   
 "Sachet water [A]" = 92,   
 "Missing" = 99),  
   
   
 drinkingWater = "Main Source of drinking water",  
 drinkingWater = c(  
 "Improved sources" = 1,  
 "Unimproved sources" = 2  
 ),  
 toiletType = "Type of sanitation facility",  
 toiletType = c(  
 "Improved" = 1,  
 "Unimproved" = 2,  
 "Open defecation (no facility, bush, field)" = 3  
 ),  
 flush = "Flush/Pour flush to:",  
 sharedToilet = c(  
 "Not shared" = 0,  
 "5 households or less" = 1,  
 "More than 5 households" = 2,  
 "Public facility" = 3,  
 "DK/Missing" = 9  
 ),  
 improvedWater = "Percentage using improved sources of drinking water [1]",  
 HH6 = c("URBAN" = 1, "RURAL" = 2)  
# ,WS1 = ""  
  
   
   
 )  
  
  
  
  
# hh <- as\_factor(hh, only\_labelled = TRUE)

# Tables

* Note - lots of understanding and formatting etc to get through on tables

explore expss package which seems to have capability to produce ctables equivalents

Tanker Truck

Cart with small tank

Bottled water [A]

Sachet water [A]

Missing

URBAN

208

912

2519

18

0

RURAL

131

306

205

0

0

## Test table packages

After lots of exploring for table producing, expss package in R seems most promising, as it seems to have direct ability to produce any ctable from SPSS

Much formatting to be done

#hh.tab %>% as\_factor(only\_labelled = TRUE) %>%   
# as\_survey(weights = c(hhweightHH48)) %>%  
# group\_by(HH6, WS1, nhhmem) %>%  
# summarize(n = survey\_total()) %>%  
# spread(WS1, n) %>%   
# kable(format = "html", digits = 3, row.names = TRUE) %>%   
# kable\_styling(bootstrap\_options = "striped", font\_size = 12, full\_width = F)  
  
  
hh.tab = as\_factor(zap\_labels(hh.tab), only\_labelled = TRUE)  
  
#CrossTable(hh.tab$HH6, hh.tab$WS1, expected = FALSE, prop.col = FALSE, prop.r = TRUE,   
# prop.t = F, prop.chisq = F, chisq = F) %>%   
# kable(format = "html", digits = 3, row.names = TRUE) %>%   
# kable\_styling(bootstrap\_options = "striped", font\_size = 12, full\_width = F)  
  
  
  
#tabl = CrossTable(hh.tab$HH6, hh.tab$WS1, expected = FALSE,   
# prop.col = FALSE, prop.r = TRUE,   
# prop.t = F, prop.chisq = F, chisq = F, max.width = 2)[[2]]  
  
#tabl %>% kable(format = "html", digits = 3, row.names = TRUE) %>%   
# kable\_styling(bootstrap\_options = "striped", font\_size = 12, full\_width = F)  
  
calculate(hh.tab, cro\_rpct(HH6, list(total(), WS1),   
 total\_statistic = "u\_cases")) %>%   
 kable(format = "html", digits = 3, row.names = TRUE) %>%   
 kable\_styling(bootstrap\_options = "striped", font\_size = 12, full\_width = F)

row\_labels

#Total

Main source of drinking water|Tanker Truck

Main source of drinking water|Cart with small tank

Main source of drinking water|Bottled water [A]

Main source of drinking water|Sachet water [A]

Main source of drinking water|Missing

1

Area|URBAN

100

5.688

24.938

68.882

0.492

NA

2

Area|RURAL

100

20.405

47.664

31.931

NA

NA

3

Area|#Total cases

4299

339.000

1218.000

2724.000

18.000

NA

# 10.1.2

Recode WS4

Odd that 1.2 syntax file has variable label for drinkingWater exact same as time var label….

hh <- hh %>%   
 mutate(  
 time = case\_when(  
 WS4 %in% 0:30 ~ 2,  
 WS4 %in% 31:990 ~ 3,  
 WS4 %in% 998:999 ~ 9,  
 WS1 %in% 11:13 | WS2 %in% 11:13 | WS3 %in% 1:2 ~ 1  
 )) %>%   
 mutate(INDWS2 = 0) %>%   
 mutate(INDWS2 = case\_when(  
 WS1 %in% c(11, 12) ~ 100,  
 WS1 %in% c(61, 71, 72) & WS4 <= 30 ~ 100,  
 WS1 %in% c(13, 14, 21, 31, 41, 51) & WS3 %in% 1:2 | WS4 <= 30 ~ 100,  
 WS1 %in% 91:92 & WS2 %in% 11:12 ~ 100,  
 WS1 %in% 91:92 & (WS2 %in% c(61, 71, 72) & WS4 <= 30) ~ 100,  
 WS1 %in% 91:92 & WS2 %in% c(13, 14, 21, 31, 41, 51) &   
 (WS3 %in% 1:2 | WS4 <= 30) ~ 100  
 )  
   
 )

## 10.1.2 Labels

hh <- hh %>%   
 apply\_labels(  
 time = "Time to source of drinking water",  
 time = c(  
 "Water on premises" = 1,  
 "Up to and including 30 minutes [A]" = 2,  
 "More than 30 minutes" = 3,  
 "DK/Missing" = 9  
 ),  
 drinkingWater = "Time to source of drinking water",  
 drinkingWater = c(  
 "Users of improved drinking water sources" = 1,  
 "Users of unimproved drinking water sources" = 2  
 ),  
 INDWS2 = "Percentage using basic drinking water services [1]"  
   
 )

## 10.1.2 Tables

ctables  
 /vlabels variables = time  
 display = none  
 /table total [c]  
 + HH6 [c]  
 + HH7 [c]  
 + helevel [c]  
 + ethnicity [c]  
 + windex5 [c]  
 by  
 drinkingWater [c] >  
 time[c][layerrowpct.validn '' f5.1]  
 + total100 [s] [mean '' f5.1]  
 + INDWS2 [s] [mean '' f5.1]  
 + nhhmem [s] [count '' f5.0]  
 /categories variables=all empty=exclude missing=exclude