\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* CREATING an MPI IN STATA

\* OPHI Training 2018

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* Clear information in the memory

clear all

set more off

\* Change working folder

cd "C:\Dropbox\Maldives\Presentations\Example Stata"

\* Save a record of everything

capture log close

log using log\_stata.log, replace

\* Open a dataset

use "pak\_dhs12-13\_sample.dta", clear

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* BUILDING THE DEPRIVATION MATRIX

\* Each vector provides information about the condition of deprivation of each

\* individual in an specific indicator. The mean of this vector shows the

\* incidence of each deprivation on the total population

\* (Uncensored Headcount Ratios)

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* WATER

\* -----------------------------------------------------------------------------

/\*

A household is non deprived if: it has piped water, public tap, borehole,

hand pump, protected well, protected spring, rainwater, filtration plant,

or bottled water

A household is deprived if: it gets water from an unprotected well,

unprotected spring; river/dam/stream/pond/canal; tanker truck,

cart with small tank, other

\*/

lookfor water

codebook water, tab(20)

recode water (11/31=0)(32=1)(41=0)(42/43=1)(51=0)(61/62=1)(63/73=0)(96=1)(99=.), gen(hh\_d\_water)

lab var hh\_d\_water "Household deprived in Access to Safe Water"

\* -----------------------------------------------------------------------------

\* ASSETS

\* -----------------------------------------------------------------------------

\* A household is deprived if it has less than 2 small assets and no car

egen n\_assets = rowtotal(radio television refrigerator bicycle motorbike), missing

tab n\_assets [aw = weight], miss

gen hh\_d\_assets = (n\_assets<2) if n\_assets!=.

tab n\_assets hh\_d\_assets [aw = weight]

\* Using the car as veto

replace hh\_d\_assets = 0 if car==1

lab var hh\_d\_assets "Household deprived in Assets"

table n\_assets car hh\_d\_assets [aw = weight]

\* -----------------------------------------------------------------------------

\* SCHOOLING

\* -----------------------------------------------------------------------------

\* A household is deprived if no member older than 15 has completed 5+ years of schooling

gen d\_scho = (eduyears<5) if age>15 & age!=. & eduyears!=.

bys hh\_id: egen hh\_d\_school = min(d\_scho)

lab var hh\_d\_school "Household deprived in Years of Schooling"

\* -----------------------------------------------------------------------------

\* NUTRITION

\* -----------------------------------------------------------------------------

\* A household is deprived if any child under 5 with nutritional information is undernourished

sum z\_scorewa [aw = weight]

gen d\_nutrition = (z\_scorewa<-2) if z\_scorewa!=.

tab d\_nutrition [aw = weight], miss

bys hh\_id: egen hh\_d\_nutri = max(d\_nutrition)

replace hh\_d\_nutri = 0 if no\_child\_eligible==1

lab var hh\_d\_nutri "Household deprived in Nutrition"

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* RELEVANT SAMPLE

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* We construct a filter variable that identifies the observations with info for all relevant indicators

gen sample = (hh\_d\_water~=. & hh\_d\_assets~=. & hh\_d\_school~=. & hh\_d\_nutri~=.)

sum hh\_d\_water hh\_d\_assets hh\_d\_school hh\_d\_nutri [aw = weight] if sample==1

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* ANALYSIS OF THE MPI INDICATORS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\*\*\* 1. MISSING VALUES

\* Final check to see the total number of missing values we have for each variable

\* Variables should not have high proportion of missing values at this stage

\* The command might need to be installed: write "findit mdesc" in the command window, and install it

mdesc hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets

\*\*\* We keep only those observations with information for all relevant indicators and that are usual members of the household

keep if sample==1 & hv102==1

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* UNCENSORED HEADCOUNT RATIOS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

sum hh\_d\_school [aw = weight]

gen new\_uncen\_H\_temp = r(mean)\*100

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

sum `var' [aw = weight]

gen uncen\_H\_`var' = r(mean)\*100

lab var uncen\_H\_`var' "Uncensored Headcount Ratio: Percentage of people who are deprived in …"

}

\*

save "pak\_dhs2012-13\_cleaned.dta", replace

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* SETTING WEIGHTS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* Define vector 'w' of weights

\* Change according to your specification. Remember the sum of weights MUST be

\* equal to 1 or 100%

foreach var in hh\_d\_school hh\_d\_nutri {

gen w\_`var' = 1/3

lab var w\_`var' "Weight `var'"

}

\*

foreach var in hh\_d\_water hh\_d\_assets {

gen w\_`var' = 1/6

lab var w\_`var' "Weight `var'"

}

\*

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* WEIGTHED DEPRIVATION MATRIX

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* The following commands multiply the deprivation matrix by the weight of each

\* indicator.

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

gen g0\_w\_`var' = `var' \* w\_`var'

lab var g0\_w\_`var' "Weigthed Deprivation of `var'"

}

\*

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* COUNTING VECTOR

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* Generate the vector of individual weighted deprivation score, 'c'

egen c\_vector = rowtotal(g0\_w\_\*)

lab var c\_vector "Counting Vector"

tab c\_vector [aw = weight], m

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* INDENTIFICATION

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* Using different poverty cut-offs (i.e. different k)

forvalue k = 10(10)100 {

gen multid\_poor\_`k' = (c\_vector >= `k'/100)

lab var multid\_poor\_`k' "Poverty Identification with k=`k'%"

}

\*

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* CENSORED COUNTING VECTOR

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* Generate the censored vector of individual weighted deprivation score, 'c(k)',

\* providing a score of zero if a person is not poor

forvalue k = 10(10)100 {

gen cens\_c\_vector\_`k' = c\_vector

replace cens\_c\_vector\_`k' = 0 if multid\_poor\_`k'==0

}

\*

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* M0, H and A for all the possible cutoffs so far

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* By sumarizing (obtaining the mean) of the identification vector, the individual deprivation share,

\* and the individual censored c vector at any level of k we will obtain the Multidimensional Headcount

\* Ratio (H), the Intensity of Poverty among the Poor (A), and the Adjusted Headcount Ratio (M0), respectively.

\*\*\* H \*\*\*

sum multid\_poor\_\* [aw = weight], sep(15)

\*\*\* A \*\*\*

forvalue k = 10(10)100 {

sum cens\_c\_vector\_`k' if multid\_poor\_`k'==1 [aw = weight], sep(15)

}

\*\*\* MPI \*\*\*

forvalue k = 10(10)100 {

sum cens\_c\_vector\_`k' [aw = weight], sep(15)

}

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* M0, H and A for k = 40%

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

/\*( NOW WE CHOOSE A VALUE OF k )\*/

local k = 40

\* -----------------------------------------------------------------------------

\* CENSORED DEPRIVATION MATRIX

\* -----------------------------------------------------------------------------

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

gen g0\_`k'\_`var' = `var'

replace g0\_`k'\_`var' = 0 if multid\_poor\_`k'==0

}

\* -----------------------------------------------------------------------------

\* HEADCOUNT/INCIDENCE OF MULTIDIMENSIONAL POVERTY FOR k = 40% (H)

\* -----------------------------------------------------------------------------

sum multid\_poor\_`k' [aw = weight]

gen H = r(mean)\*100

lab var H "Headcount Ratio (H): % Population in multidimensional poverty"

\* -----------------------------------------------------------------------------

\* INTENSITY OF POVERTY AMONG THE POOR FOR k = 40% (A)

\* -----------------------------------------------------------------------------

sum cens\_c\_vector\_`k' [aw = weight] if multid\_poor\_`k'==1

gen A = r(mean)\*100

lab var A "Intensity of deprivation among the poor (A): Average % of weighted deprivations"

\* -----------------------------------------------------------------------------

\* ADJUSTED HEADCOUNT RATIO (M0) FOR k = 40%

\* -----------------------------------------------------------------------------

sum cens\_c\_vector\_`k' [aw = weight]

gen M0 = r(mean)

lab var M0 "Adjusted Headcount Ratio (M0 = H\*A): Range 0 to 1"

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* CENSORED HEADCOUNT RATIOS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* The Censored Headcount Ratio of an indicator is the proportion of the population

\* that are poor AND deprived in that indicator.

\* They can be obtained as the mean of each column of the censored deprivation matrix

local k = 40

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

sum g0\_`k'\_`var' [aw = weight]

gen cen\_H\_`var' = r(mean)\*100

lab var cen\_H\_`var' "Censored Headcount Ratio: % of people who are poor and deprived in …"

}

fsum uncen\_H\_\* cen\_H\_\* [aw = weight]

sum cen\_H\_\*

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* DIMENSIONAL BREAKDDOWN: PERCENTAGE CONRIBUTIONS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

gen perc\_cont\_`var' = (cen\_H\_`var' \* w\_`var') / M0

lab var perc\_cont\_`var' "Percentage contribution to M0"

}

sum perc\_cont\_\* [aw = weight], sep(15)

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* SUBGROUP DECOMPOSITION

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

local k = 40

\* Uncensored Headcount Ratios by region

tabstat hh\_d\_\* [aw = weight], by(region)

local k = 40

\* Incidence of Poverty (H) by region

tabstat multid\_poor\_`k' [aw = weight], by(region)

\* Intensity of Poverty (A) by region

tabstat cens\_c\_vector\_`k' [aw = weight] if multid\_poor\_`k'==1, by(region)

\* Adjusted Headcount Ratio (M0) by region

tabstat cens\_c\_vector\_`k' [aw = weight], by(region)

local k=40

\* Censored Headcount Ratios by region

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

forvalue r = 1/6 {

sum g0\_`k'\_`var' [aw = weight] if region==`r'

gen cen\_H\_r`r'\_`var' = r(mean)\*100

lab var cen\_H\_r`r'\_`var' "Censored Headcount Ratio - region `r'"

}

}

sum cen\_H\_r\* [aw = weight], sep(6)

\* Contributions by region

forvalue r = 1/6 {

foreach var in hh\_d\_school hh\_d\_nutri hh\_d\_water hh\_d\_assets {

sum cens\_c\_vector\_`k' [aw = weight] if region==`r'

loc M0\_r`r' = r(mean)

gen perc\_cont\_r`r'\_`var' = (cen\_H\_r`r'\_`var' \* w\_`var') / `M0\_r`r''

lab var perc\_cont\_r`r'\_`var' "Percentage contribution to M0 - region `r'"

}

}

sum perc\_cont\_r\* [aw = weight], sep(7)

save "MyFirstMPI.dta", replace

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* COLLAPSE RESULTS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

forvalues k=10(10)100{

gen A\_`k' = cens\_c\_vector\_`k' if multid\_poor\_`k'==1

}

collapse A\_\* uncen\_H\_\* w\_\* c\_vector multid\_poor\_\* cens\_c\_vector\_\* cen\_H\_\* perc\_cont\_\* [aw = weight], by(region)

save "Collapsed\_results by region.dta", replace

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* POVERTY MAPS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

/\*

1. To obtain the shapefiles go to http:\*www.diva-gis.org/gdata and download the data for your country.

Then, se the following command to transform the shapefile in a stata file:

shp2dta using XXX\_adm1, database(region) coordinates(map) genid(id) gencentroids(center)

In the new dataset called "region" check the id for each region (br id NAME\_1 VARNAME\_1).

In the collapsed results dataset generate a new variable "id" following the structure of the dataset

"regions.dta"

\*/

cd "C:\TanzaniaHBS2012\sdr\_subnational\_boundaries\_2018-05-11\shps"

ssc install shp2dta

ssc install spmap

shp2dta using sdr\_subnational\_boundaries, database(tz\_database) coordinates(tz\_coord)

\* -----------------------------------------------------------------------------

\* INCIDENCE OF POVERTY (H)

\* -----------------------------------------------------------------------------

ta multid\_poor\_40

cd "C:\Dropbox\Maldives\Presentations\Example Stata"

spmap multid\_poor\_40 using Pakistan\_coord.dta, id(region) clmethod(unique) fcolor(Reds) ///

legend(pos(5) subtitle("Headcount Ratio", size(vsmall))) ///

title("Incidence of Multidimensional Poverty H" "in Pakistan, k=40%")

gr export Incidence\_Pakistan.emf, replace

\* -----------------------------------------------------------------------------

\* MULTIDIMENSIONAL POVERTY INDEX (M0)

\* -----------------------------------------------------------------------------

ta cens\_c\_vector\_40

spmap cens\_c\_vector\_40 using pak\_c.dta, id(region) clmethod(unique) fcolor(Reds) ///

legend(pos(5) subtitle("M0", size(vsmall))) line(data("pak\_c.dta")) ///

title("Multidimensional Poverty in Pakistan (MPI), k = 40%")

gr export M0\_Pakistan.emf, replace

\*/

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* ROBUSTNESS, SENSITIVITY AND STANDARD ERRORS

\* -----------------------------------------------------------------------------

\* -----------------------------------------------------------------------------

\* We open the clean dataset

clear

use "MyFirstMPI.dta"

\* -----------------------------------------------------------------------------

\* RANK ROBUSTNESS COMPARISONS

\* -----------------------------------------------------------------------------

\* Robustness tests are based on the coefficient of rank correlations Kendall tau-b, which

\* measures the association between pairs, given the position that each takes when results are

\* sorted using different poverty indices.

\* These different poverty indices can be obtained changing the weights of indicators or the

\* povetry cut-off (k).

\* a. Variations in weights: several MPI are computed keeping dimensions/indicators and deprivations

\* cut-offs unchanged; only the weights are modified. Once all the MPI have been computed, figures by

\* subnational regions can be obtained and regions ranked. The Kendall tau-b coefficient can then be

\* computed over the rankings.

\* b. Variations in the poverty cut-offs (k): several MPI are computed keeping the structure unchanged

\* and also adjusting the k-value. Once all the MPI have been computed, figures by subnational regions

\* can be obtained and regions ranked. The Kendall tau-b coefficient can then be

\* computed over the rankings.

forvalues k = 10(10)100 {

gen H\_`k' = .

forvalues r = 1/7 {

sum multid\_poor\_`k' [aw = weight] if region==`r'

replace H\_`k' = r(mean)\*100 if region==`r'

}

}

ktau H\_10 H\_20 H\_30 H\_40 H\_50 H\_60 H\_70 H\_80 H\_90 H\_100, stats(taub score se p)

\* -----------------------------------------------------------------------------

\* STANDARD ERRORS

\* -----------------------------------------------------------------------------

\* We open the clean dataset

clear

use "MyFirstMPI.dta"

\* Set the characteristics of the survey

svyset psu [pw = weight], strata(strata)

\* Incidence of Poverty (H)

\* For details and discussions see equations (8.13) and (8.31), chapter 8 of OPHI book

forvalue k = 10(10)100 {

svy: mean multid\_poor\_`k'

gen se\_H\_`k' = (\_se[multid\_poor\_`k'])

gen lb\_H\_`k' = \_b[multid\_poor\_`k'] - 1.96 \* se\_H\_`k'

gen ub\_H\_`k' = \_b[multid\_poor\_`k'] + 1.96 \* se\_H\_`k'

}

sum multid\_poor\_\* lb\_H\_\* ub\_H\_\* [aw = weight]

\* Adjusted Headcount Ratio (M0)

\* For details and discussions see equations (8.11) and (8.30), chapter 8 of OPHI book

forvalue k = 10(10)100 {

svy: mean cens\_c\_vector\_`k'

gen se\_M0\_`k' = (\_se[cens\_c\_vector\_`k'])

gen lb\_M0\_`k' = \_b[cens\_c\_vector\_`k'] - 1.96 \* se\_M0\_`k'

gen ub\_M0\_`k' = \_b[cens\_c\_vector\_`k'] + 1.96 \* se\_M0\_`k'

}

sum cens\_c\_vector\_\* lb\_M0\_\* ub\_M0\_\* [aw = weight]

\* Average Deprivation among the Poor (A)

\* For details and discussions see equations (8.19), (8.35) and (8.36), chapter 8 of OPHI book

forvalue k = 10(10)100 {

svy: mean multid\_poor\_`k' cens\_c\_vector\_`k'

mat cov = e(V)

loc cov = cov[2,1]

loc var\_H = cov[1,1]

loc var\_MPI = cov[2,2]

gen se\_A\_`k' = ((`var\_MPI'/\_b[multid\_poor\_`k']^2) + (((\_b[cens\_c\_vector\_`k']/\_b[multid\_poor\_`k']^2)^2)\*(`var\_H')) ///

- 2\*((\_b[cens\_c\_vector\_`k']/\_b[multid\_poor\_`k']^3)\*`cov'))^0.5

gen lb\_A\_`k' = (\_b[cens\_c\_vector\_`k']/\_b[multid\_poor\_`k']) - 1.96 \* se\_A\_`k'

gen ub\_A\_`k' = (\_b[cens\_c\_vector\_`k']/\_b[multid\_poor\_`k']) + 1.96 \* se\_A\_`k'

}

\* -----------------------------------------------------------------------------

\* DOMINANCE AMONG SUBNATIONAL REGIONS

\* -----------------------------------------------------------------------------

\* For M0

collapse mean cens\_c\_vector\_10 cens\_c\_vector\_20 cens\_c\_vector\_30 cens\_c\_vector\_40 ///

cens\_c\_vector\_50 cens\_c\_vector\_60 cens\_c\_vector\_70 cens\_c\_vector\_80 ///

cens\_c\_vector\_90 cens\_c\_vector\_100 [aw = weight], by(region)

reshape long cens\_c\_vector\_, i(region) j(k)

gen cens\_c\_vector\_1 = cens\_c\_vector\_ if region==1

label var cens\_c\_vector\_1 "Balochistan"

gen cens\_c\_vector\_2 = cens\_c\_vector\_ if region==2

label var cens\_c\_vector\_2 "Islamabad (ICT)"

gen cens\_c\_vector\_3 = cens\_c\_vector\_ if region==3

label var cens\_c\_vector\_3 "Khyber Pakhtunkhawa"

gen cens\_c\_vector\_4 = cens\_c\_vector\_ if region==4

label var cens\_c\_vector\_4 "Gilgit Baltistan"

gen cens\_c\_vector\_5 = cens\_c\_vector\_ if region==5

label var cens\_c\_vector\_5 "Punjab"

gen cens\_c\_vector\_6 = cens\_c\_vector\_ if region==6

label var cens\_c\_vector\_6 "Sindh"

graph twoway line cens\_c\_vector\_1 k || line cens\_c\_vector\_2 k || line cens\_c\_vector\_3 k || line cens\_c\_vector\_4 k || line cens\_c\_vector\_5 k || line cens\_c\_vector\_6 k

\* For H

clear

use "MyFirstMPI.dta"

collapse mean multid\_poor\_10 multid\_poor\_20 multid\_poor\_30 multid\_poor\_40 multid\_poor\_50 multid\_poor\_60 ///

multid\_poor\_70 multid\_poor\_80 multid\_poor\_90 multid\_poor\_100 [aw = weight], by(region)

reshape long multid\_poor\_, i(region) j(k)

gen multid\_poor\_1 = multid\_poor\_ if region==1

label var multid\_poor\_1 "Balochistan"

gen multid\_poor\_2 = multid\_poor\_ if region==2

label var multid\_poor\_2 "Islamabad (ICT)"

gen multid\_poor\_3 = multid\_poor\_ if region==3

label var multid\_poor\_3 "Khyber Pakhtunkhawa"

gen multid\_poor\_4 = multid\_poor\_ if region==4

label var multid\_poor\_4 "Gilgit Baltistan"

gen multid\_poor\_5 = multid\_poor\_ if region==5

label var multid\_poor\_5 "Punjab"

gen multid\_poor\_6 = multid\_poor\_ if region==6

label var multid\_poor\_6 "Sindh"

graph twoway line multid\_poor\_1 k || line multid\_poor\_2 k || line multid\_poor\_3 k || line multid\_poor\_4 k || line multid\_poor\_5 k || line multid\_poor\_6 k

\* -----------------------------------------------------------------------------

\* TEST OF DIFFERENCE BETWEEN REGIONS (e.g. Balochistan and Sindh, M0)

\* -----------------------------------------------------------------------------

clear

use "MyFirstMPI.dta"

svyset psu [pw = weight], strata(strata)

svy: mean cens\_c\_vector\_30, over(region)

test \_b[Balochistan] = \_b[Sindh]

\* -----------------------------------------------------------------------------

\* NOTES

\* -----------------------------------------------------------------------------

\* For discussions regarding statistical tests, please see Sections 8.2.2 and 8.2.3 of OPHI Book

\* For Robustness and dominance analysis with Statistical inference, see Section 8.3 of OPHI Book.

\* For those who are interested in bootstrap, see the Appendix of Chapter 8 of the OPHI Book.