Generating descriptive statistics from the MXFLS

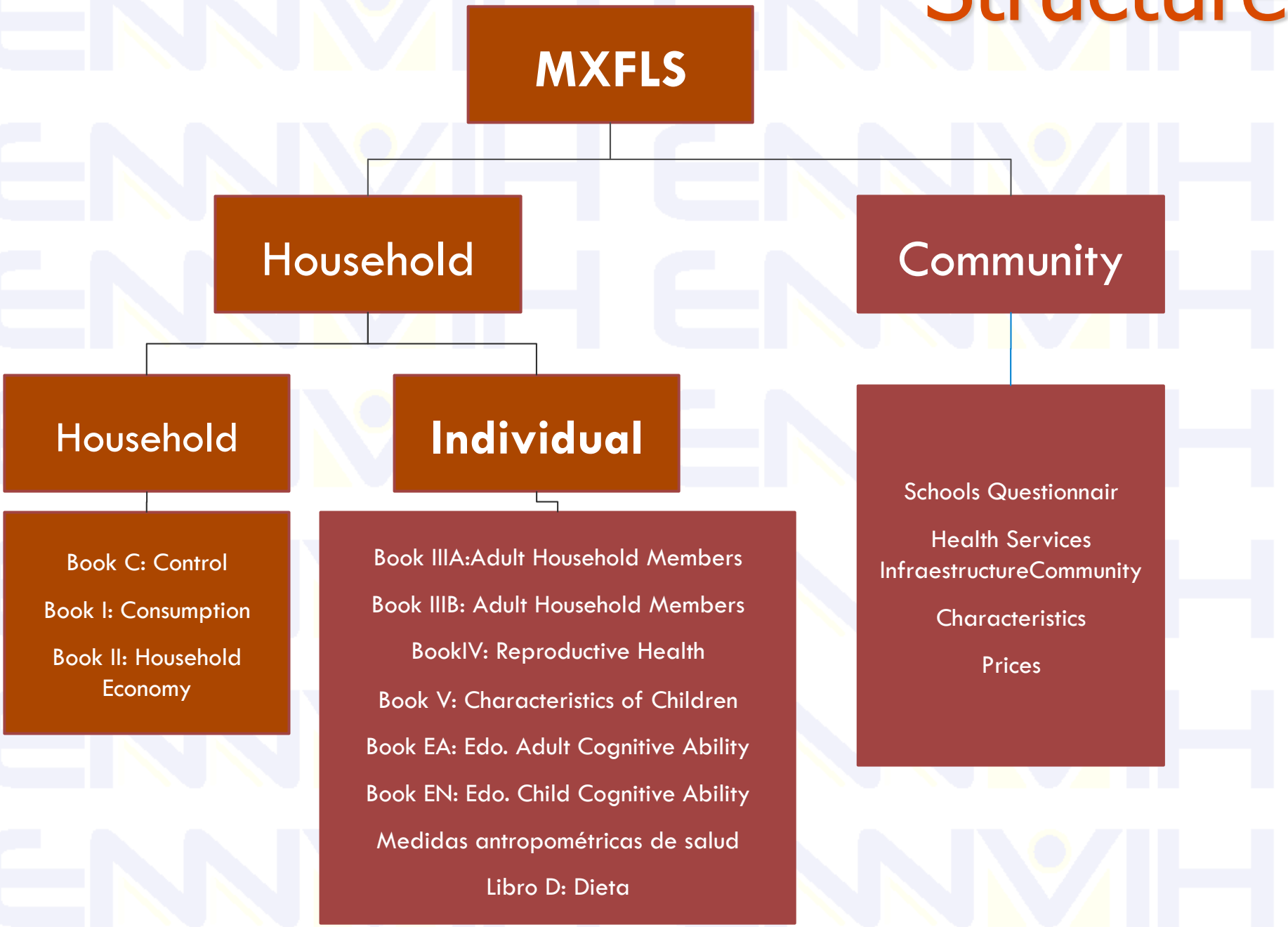
Alicia Santana

Characteristics

- Multi-thematic and longitudinal database
- National, urban-rural and regional representation
- The approximate sampling size is 8,440 households with approximately 35,000 individual interviews

Panel 2002 ⇢ 2005 ⇢ 2009 ⇢

Structure


MXFLS
Household
Community
Household
Individual
Book C: Control
Book I: Consumption
Book II: Household Economy
Book IIIA:Adult Household Members
Book IIIB: Adult Household Members
BookIV: Reproductive Health
Book V: Characteristics of Children
Book EA: Edo. Adult Cognitive Ability
Book EN: Edo. Child Cognitive Ability
Medidas antropométricas de salud
Libro D: Dieta
Schools Questionnair
Health Services InfraestructureCommunity
Characteristics
Prices

Topics

Household

- Dwelling characteristics
- Household economy
- Land
  INR- Rural income
- Household Assets
- Household economic shocks

Individuals

Education
Schooling interruptions
Individual non-labor income
Marital history
Migration
Employment
Adult time allocation
Individual crime and victimization
Health condition
Emotional well-being
Insurance condition
Credit
Transfers
Pregnancy history
Contraception
Health Measures

Community

Schools Questionnair

Health Services
InfraestructureCommunity

Characteristics

Prices

http://www.ennvih-mxfls.org/

Longitudinal database

Objective: The purpose of MxFLS-2 is to re-interview all individuals and households who were interviewed during the base line (MxFLS-1) and all individuals and households who, due to the growth and family developments from our original sample size, were added on.

Attritions: For MXFLS it was decided refresh our sample size naturally. In this way, a home or individual is defined as panel if they belong to the original sample size and if they were interviewed during 2002. If on the other hand, new individuals who were integrated into the home and were not interviewed in 2002, but are now considered family members, they would be classified as new members and interviewed for the first time.

Combining files

Suppose your objective is to analyze a file containing information about different books, you need organize each of the two or more files by order of folio and Ls and subsequently we merge by folio and Ls

Combining MxFLS-1 information with MxFLS-2

Panel Households folio is the same for MxFLS1 and MXFLS2, and the last two digits will be 00.

Combining files

New Home Folio: If an MxFLS-1 original member has departed from the home and forms his/her own household, that individual's folio will be made up of the first six digits stemming from the original home and the last two digits will be made up of the individual identifier in MxFLS-1.

pid_link: In order to have an individual identifier in both databases, the identifier pid_link was created, allowing individual identification in both survey rounds. The pid_link is comprised of the individuals' original folio and ls identifiers, that is to say, the folio and Ls corresponding to MxFLS-1, regardless if the individual is in the original or new household during MxFLS-2. The pid_link is made up of 10 digits, the first 8 digits are the folio digits and the last two correspond to the LS.

Combining files

In STATA, the way to create the original folio from the pid_link is as follows

gen str8 var = substr (pid_link, 1,8)

In contrast to MxFLS-2, the folio from the first survey round is numeric and in order to carryout the adding and combination of both databases, it is necessary to have the same format for both folios. To turn the MxFLS-1 folio to STRING using STATA, you could use the following instruction

gen str8 var1=string(folio, "%08.0f")

STATA commands in order to create MxFLS-1 pid_links:

gen str8 var1 = string (folio, "%08.0f")

gen str2 var2 = string (ls, "%02.0f")

gen pid_link = var1 + var2

Weights

The weight is the value used to adjust the information of the variables captured through a sample in order to multiply the observations to better represent the whole population.

In order to adjust the basic weights, we had to consider the non-response rate of each book independently, such is the reason why each book has a different set of weights

Weights

2005 Weights:

a) Continue to expand through 2002, so you must use the 2002 weights, adjusting for nonresponse and adjusting for demographic factor to expand the population. These weights are called longitudinal weights.

b) Expand the population of 2005. To do this we use the weights of 2002, adjusted for nonresponse and adjust for the demographic factor that makes the survey representative of 2005.

Aplication

Migration in Mexico

Estadísticas descriptivas

by migration: tabstat añosesc edad02 if edad02>=15, stats(mean median sd var min max)

Characteristics of the migrant population at the time of departure			
Years of schooling	**non-migrant**	**Internal**	**US**
mean	7.3636	9.0505	7.7907
p50	7.0000	9.0000	8.0000
sd	4.6863	4.2310	3.4357
variance	21.9617	17.9015	11.8043
min	0.0000	0.0000	0.0000
max	18.0000	18.0000	16.0000
Age			
mean	38.0237	28.8464	25.0690
p50	35.0000	25.0000	21.0000
sd	16.9016	12.3975	10.9939
variance	285.6637	153.6990	120.8649
min	15.0000	15.0000	15.0000
max	98.0000	92.0000	77.0000

```
. mprobit mig2 edad02  edocivil2 genero  añosesc rururb pobre desempleado casa
> ahorros, base(0)

Iteration 0:   log likelihood = -5589.0881
Iteration 1:   log likelihood = -5549.1463
Iteration 2:   log likelihood = -5546.2357
Iteration 3:   log likelihood = -5546.2289
Iteration 4:   log likelihood = -5546.2289

Multinomial probit regression                   Number of obs   =      19688
                                                Wald chi2(18)   =     769.83
Log likelihood = -5546.2289                     Prob > chi2     =     0.0000
```

mig2	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_outcome_2						
edad02	-.026771	.0019767	-13.54	0.000	-.0306452	-.0228968
edocivil2	.0069851	.0548054	0.13	0.899	-.1004315	.1144017
genero	.0182726	.0450426	0.41	0.685	-.0700092	.1065544
añosesc	.0343012	.005957	5.76	0.000	.0226258	.0459767
rururb	-.0080512	.0487163	-0.17	0.869	-.1035334	.0874311
pobre	-.0393422	.0560356	-0.70	0.483	-.1491701	.0704856
desempleado	-.3784337	.0769324	-4.92	0.000	-.5292185	-.227649
casa	-.3269157	.0687357	-4.76	0.000	-.4616353	-.1921961
ahorros	.3348203	.105434	3.18	0.001	.1281734	.5414673
_cons	-1.621533	.1018426	-15.92	0.000	-1.821141	-1.421925
_outcome_3						
edad02	-.0394108	.0027055	-14.57	0.000	-.0447135	-.0341081
edocivil2	-.1364025	.0711947	-1.92	0.055	-.2759416	.0031367
genero	-.3875056	.0559665	-6.92	0.000	-.497198	-.2778132
añosesc	-.0091584	.0081427	-1.12	0.261	-.0251179	.0068011
rururb	.3309893	.0590268	5.61	0.000	.2152989	.4466798
pobre	.4298464	.0608867	7.06	0.000	.3105107	.5491821
desempleado	-.4313687	.0906552	-4.76	0.000	-.6090497	-.2536877
casa	-.1660828	.0883047	-1.88	0.060	-.3391568	.0069912
ahorros	-.0957527	.1819954	-0.53	0.599	-.4524571	.2609517
_cons	-1.276753	.1268858	-10.06	0.000	-1.525444	-1.028061

```
(mig2=0 is the base outcome)

.
```

```
. mfx, predict(p outcome(1))

Marginal effects after mprobit
      y  = Pr(mig2=1) (predict, p outcome(1))
         =  .03766329
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
edad02	-.0014489	.00011	-13.54	0.000	-.001659	-.001239	37.6189
edociv~2*	.0009108	.00321	0.28	0.776	-.005376	.007198	.612251
genero*	.0025111	.00263	0.96	0.340	-.002642	.007665	.556278
añosesc	.0020695	.00035	5.96	0.000	.001389	.00275	7.15995
rururb*	-.0017012	.00285	-0.60	0.550	-.007279	.003877	.425538
pobre*	-.0040233	.00317	-1.27	0.204	-.010235	.002188	.275549
desemp~o*	-.0175323	.00306	-5.72	0.000	-.023538	-.011527	.093306
casa*	-.0169178	.00324	-5.22	0.000	-.023275	-.010561	.230343
ahorros*	.024492	.00895	2.74	0.006	.006948	.042036	.045815

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
. mfx, predict(p outcome(2))

Marginal effects after mprobit
      y  = Pr(mig2=2) (predict, p outcome(2))
         =  .01567634
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
edad02	-.0010508	.00007	-15.42	0.000	-.001184	-.000917	37.6189
edociv~2*	-.0040966	.00219	-1.87	0.061	-.008387	.000193	.612251
genero*	-.0118732	.00179	-6.63	0.000	-.015385	-.008362	.556278
añosesc	-.0003889	.00024	-1.65	0.098	-.00085	.000072	7.15995
rururb*	.0101255	.00188	5.38	0.000	.006437	.013814	.425538
pobre*	.0148975	.00243	6.14	0.000	.010141	.019654	.275549
desemp~o*	-.0088361	.00158	-5.58	0.000	-.011941	-.005731	.093306
casa*	-.0035326	.00226	-1.56	0.119	-.007969	.000904	.230343
ahorros*	-.0037921	.00429	-0.88	0.376	-.012193	.004609	.045815

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
.
```