

Guía para el análisis de la ENSANUT 2012

Martín Romero Martínez
Instituto Nacional de Salud Pública
México 2012

Objetivo

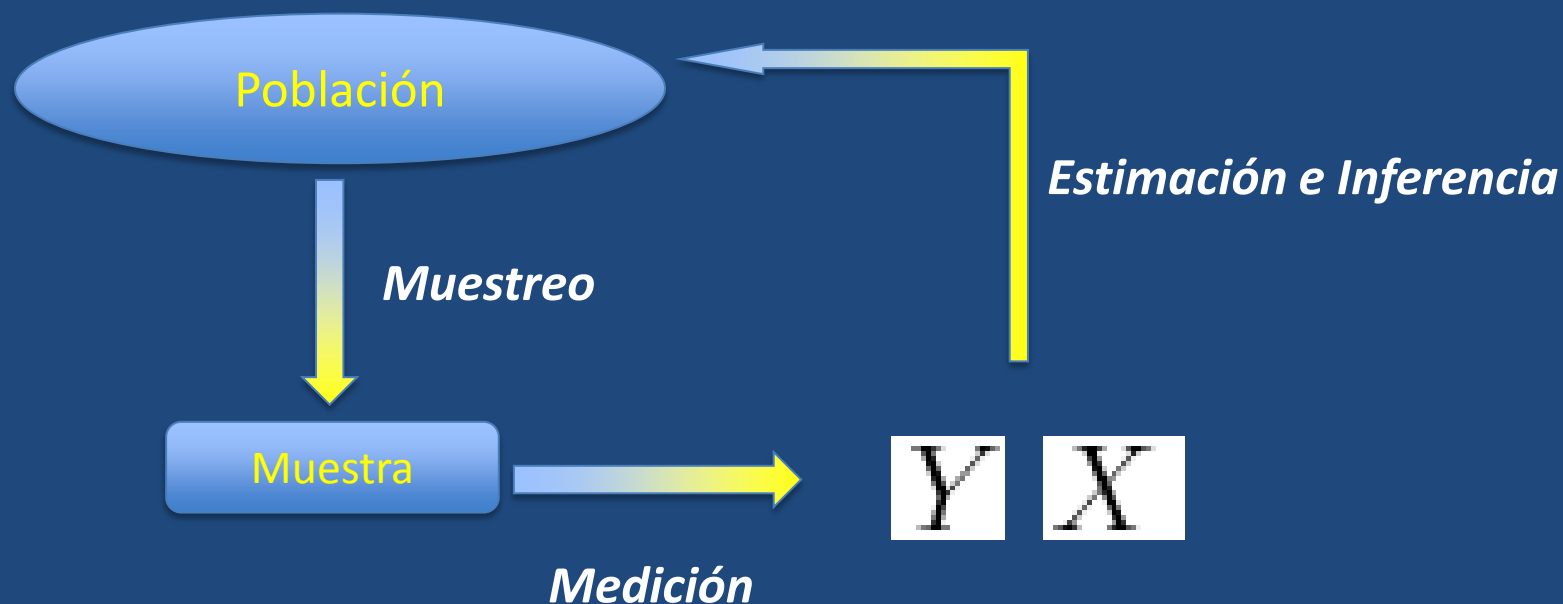
Presentar consideraciones relevantes para el análisis de la ENSANUT 2012 :

- La ENSANUT es una encuesta
- Tamaño de muestra
- El manejo de las bases de datos

La ENSANUT es una encuesta



Elementos de una encuesta



Población : Habitantes de viviendas particulares habitadas
Muestra : Elementos de la población seleccionados con probabilidades conocidas

En consecuencia ...

El análisis de la ENSANUT debe considerar los elementos del diseño :

- Muestreo
- Medición
- Inferencia



Elementos del Muestreo

- Tamaño de muestra $n = 1,712$ viviendas seleccionadas por estado
- Las unidades primarias de muestreo (UPM) son las AGEB
- Las UPM se estratificaron por dos criterios : urbanidad y marginalidad

Estratificación

Rezago Social	Urbanidad			Nueva Creación
	Rural	Urbano	Metropolitano	
Mayor	16 UPM			
Menor	32 UPM			

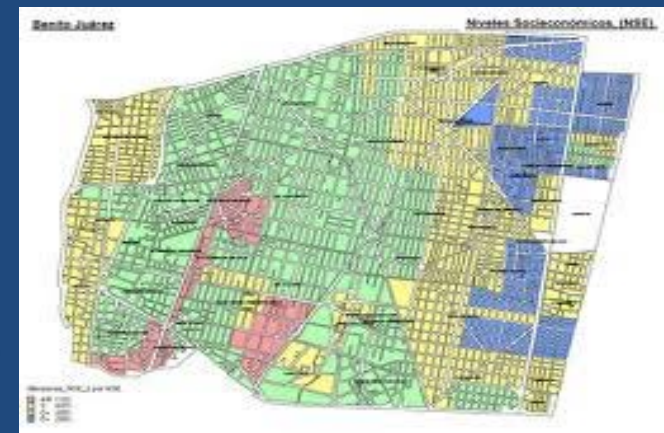
Las AGEB de mayor rezago social son las AGEB que contienen al 20% de las viviendas de un estado con mayor rezago social. El rezago social de una AGEB se definió mediante la emulación del índice de rezago social definido por el CONEVAL.

Estrato rural : localidades con menos de 2,500 habitantes

Estrato Metropolitano : localidades con mas de 100 mil habitantes, capitales de los estados, localidades no rurales de los municipios incluidos en las zonas metropolitanas definidas por INEGI en 2000

Procedimiento de selección

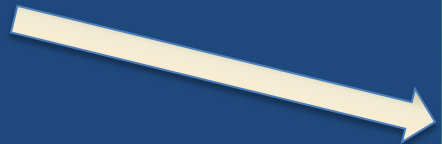
Áreas (AGEB)



Viviendas



Personas



Selección de individuos en la vivienda

Niño escolar



Adolescente



Adultos



Siempre que es posible se selecciona a :

- Un niño pre-escolar con edad 0 a 4 años
- Un niño escolar con edad 5 a 9 años
- Un adolescente con edad 10 a 19 años
- Un adulto con edad 20 años y más
- Un utilizador de servicios médicos no hospitalarios, no preventivos durante los últimos quince días

Procedimiento de medición

La información se colectó mediante la aplicación de un cuestionario, el cual se recomienda leer antes de proceder al análisis de la encuesta:

I. FACTORES DE RIESGO

0.3 ¿Has fumado por lo menos cien cigarros (5 cajetillas) de tabaco durante toda tu vida?

Sí.....

No.....

Nunca ha fumado.....

Se sugiere, contestar el cuestionario antes de analizar la encuesta.

Procedimiento de Estimación

El total de un atributo Y se estima mediante la suma ponderada :

$$\hat{Y} = \sum_i w_i Y_i$$

Individuo	Y(individuo) = 1 si tiene diabetes	W (ponderador)
A	1	10
B	0	2
C	0	3

El total de individuos con diabetes estimado es igual a = $1*10 + 0*2 + 0*3$. De modo aproximado, el ponderador i es el número de personas en la población representadas por el individuo i

Procedimiento de estimación

Las estimaciones ponderadas y no ponderadas pueden ser diferentes.

En general, es recomendable analizar los resultados ponderados de la ENSANUT porque los ponderadores contienen información sobre el diseño de muestra.

Como incorporar el diseño de muestra en el análisis (ejemplo SPSS)

CSPLAN ANALYSIS

```
/PLAN FILE= 'varianzas.csaplan'
```

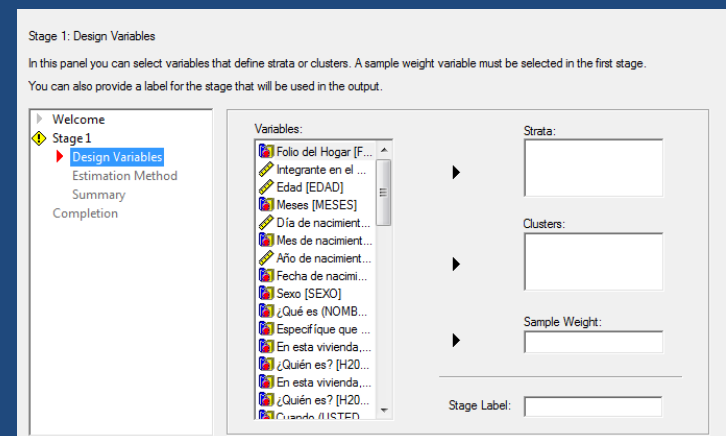
```
/PLANVARS ANALYSISWEIGHT= pondeh
```

```
/PRINT PLAN
```

```
/DESIGN STRATA= est_var CLUSTER= code_upm
```

```
/ESTIMATOR TYPE=WR.
```

	Time Series			
meri				NOMBRE) está afiliado(a) o inscrito
ing				IPM
ing				e entidad
ing				e municipio
ing				Select a Sample...
ing				Prepare for Analysis...
ing				Frequencies...
meric	1	0	Estrato	
meric	1	0	Estrato	
meric	1	0	Estrato	
meric	8	2	Estrato	



Como incorporar el diseño de muestra en el análisis (ejemplo SPSS)

CSTABULATE

/PLAN FILE = 'varianzas.csaplan'

/TABLES VARIABLES = h517

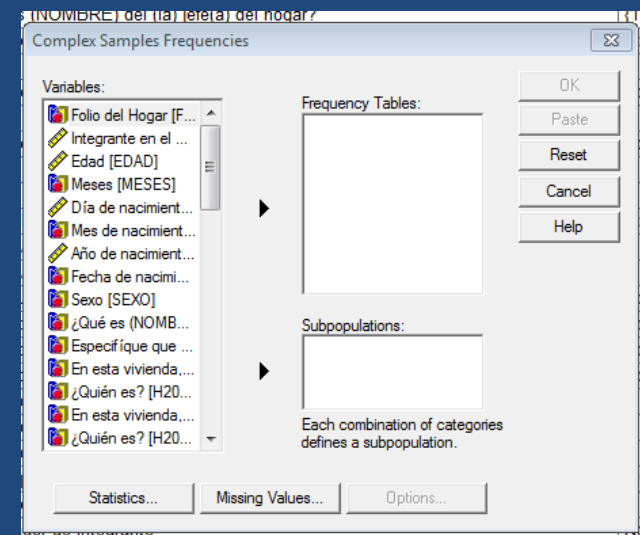
/SUBPOP TABLE = est_urb DISPLAY=LAYERED

/CELLS POPSIZE TABLEPCT

/STATISTICS CIN(95)

/MISSING SCOPE = TABLE CLASSMISSING = EXCLUDE.

Multiple response			e entidad
Missing Value Analysis...			e municipio
Complex Samples			Select a Sample...
ic	1	0	Estrato
ic	1	0	Estrato
ic	1	0	Estrato
ic	8	2	Estrato
ic	8	0	Estrato



Como incorporar el diseño de muestra en el análisis (ejemplo SPSS)

¿El dueño o propietario de la vivienda:

		Estimate	95% Confidence Interval	
			Lower	Upper
Population Size	la compró hecha?	5703583	5331371	6075794
	la mandó construir?	6827871	6547457	7108286
	la construyó el mismo?	6362444	6110116	6614773
	la construyó de otra manera?	1031277	952258.7	1110296
	Total	2E+007	2E+007	2E+007
% of Total	la compró hecha?	28.6%	26.9%	30.4%
	la mandó construir?	34.3%	33.0%	35.6%
	la construyó el mismo?	31.9%	30.7%	33.2%
	la construyó de otra manera?	5.2%	4.8%	5.6%
	Total	100.0%	100.0%	100.0%

Ejemplo en Stata

- use "\$bases_org\integrantes_p.dta", clear;
- svyset [pw= pondei],psu(code_upm) strata(est_var) singleunit(centered);
- gen afilia_1ra = h211a;
- replace afilia_1ra = 2 if afilia_1ra == 3;
- recode afilia_1ra (9=0);
- lab def afilia1 0 "Ninguna" 1 "IMSS" 2 "ISSSTE" 4 "Pemex" 5 "Defensa/Marina"
6 "Seguro Popular" 7 "Privado" 8 "Otro" 99 "NS/NR";
- lab val afilia_1ra afilia1;
- svy:prop afilia_1ra;



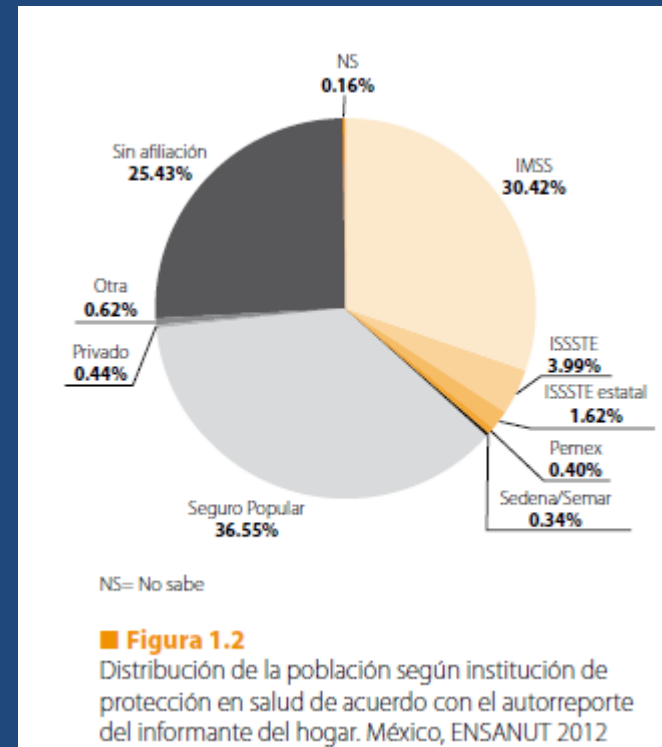
Estos son los porcentajes en la figura 1.2 del reporte

Survey: Proportion estimation

Number of strata = 155 Number of obs = 194923
 Number of PSUs = 1622 Population size = 115170278
 Design df = 1467

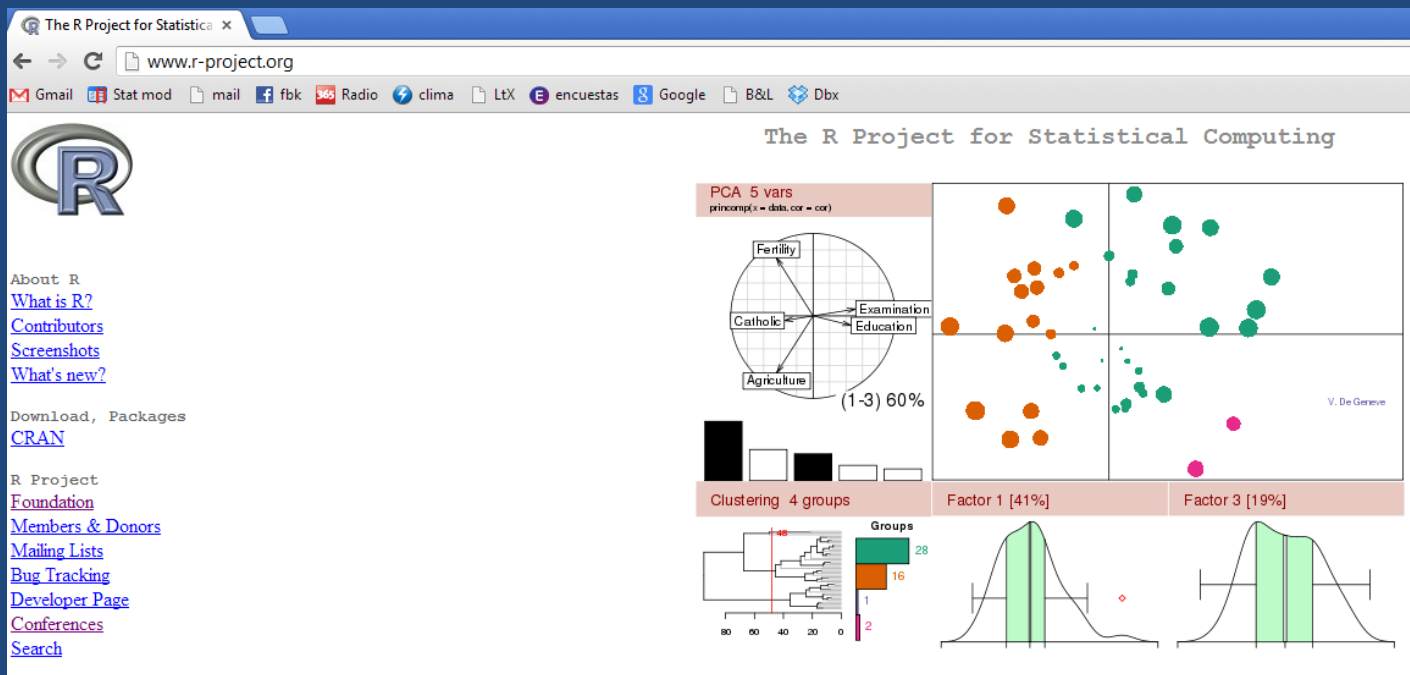
_prop_5: afilia_lra = Defensa/Marina
 _prop_6: afilia_lra = Seguro Popular
 _prop_9: afilia_lra = NS/NR

	Linearized			
	Proportion	Std. Err.	[95% Conf. Interval]	
afilia_lra				
Ninguna	.2543602	.0034141	.2476632	.2610571
IMSS	.3041183	.0046824	.2949334	.3133032
ISSSTE	.056177	.0018097	.0526271	.0597269
Pemex	.004065	.000691	.0027096	.0054205
_prop_5	.0034561	.0004774	.0025197	.0043925
_prop_6	.3655547	.0044877	.3567517	.3743577
Privado	.0044142	.0004453	.0035408	.0052877
Otro	.0062463	.0005346	.0051977	.007295
_prop_9	.0016082	.0001822	.0012507	.0019656



Existen otras alternativas

El análisis de muestras complejas puede ser realizado con igual grado de dificultad en el programa R (gratis)



Tamaño de muestra



Tamaño de muestra

La ENSANUT es una encuesta probabilista entonces es posible obtener intervalos de confianza válidos para cualquier tamaño de muestra.

Un intervalo de confianza del 95% es válido si el procedimiento que generó dicho intervalo genera un 95% de intervalos que cubren al parámetro de interés.

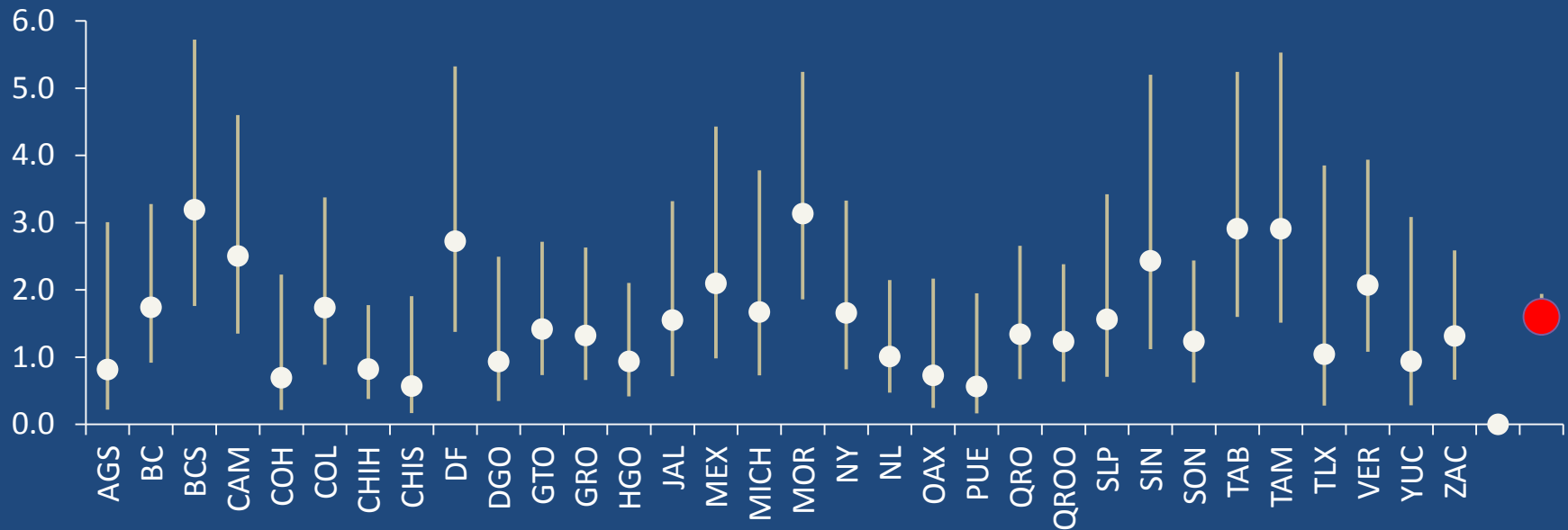
Tamaño de muestra

Cuando el tamaño de muestra es pequeño, el poder estadístico de los intervalos de confianza para la toma de decisiones decrece

Es responsabilidad del usuario el considerar el poder estadístico del tamaño de muestra

Tamaño de Muestra (Ejemplo)

Como ejemplo, la gráfica siguiente presenta las estimaciones estatales del porcentaje de adolescentes con mas de dos ayunos por semana en los últimos 3 meses (en rojo, estimación nacional)



Manejo de las bases de datos



Manejo de las bases de datos

Cada base de datos tiene un ponderador , entonces cada base de datos puede ser analizada de modo independiente. La unión de bases de datos se discute en seguida



Manejo de las bases de datos

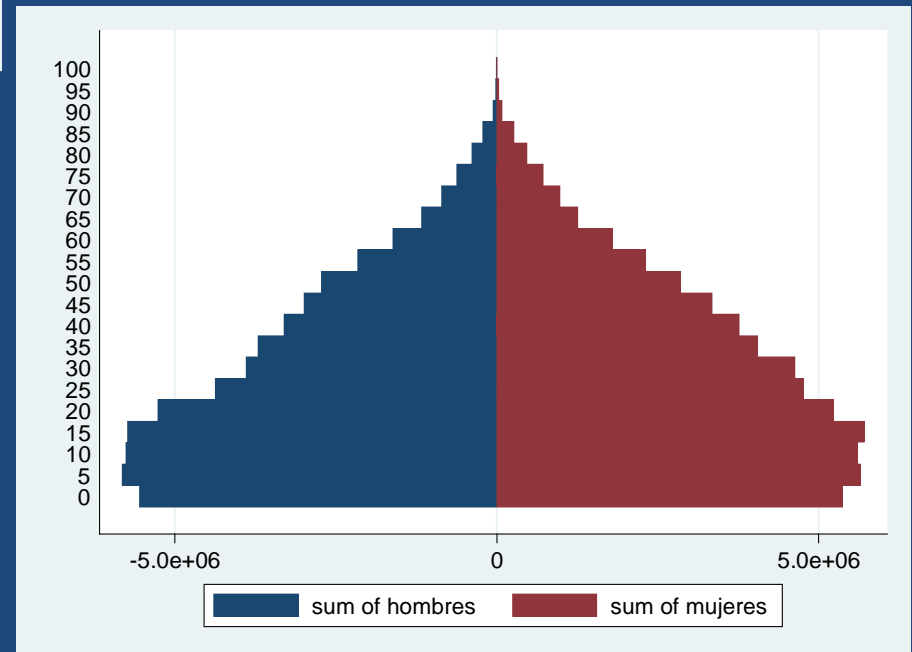
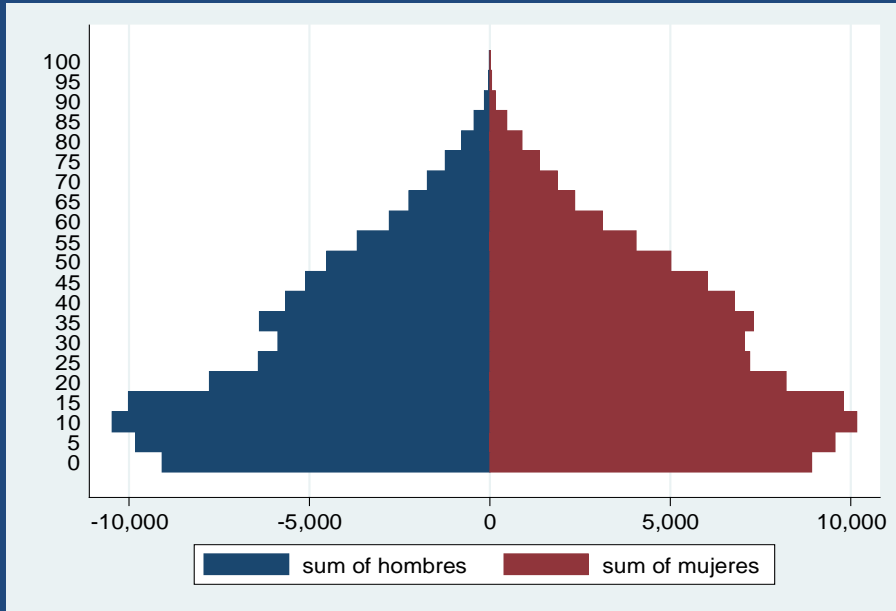
Variables incluidas en **todas** las bases de datos

Variable	Significado
Folio_c	Folio de hogar
Intp	Integrante dentro del hogar
pondef	Ponderador final
Est_var	Estrato para calcular varianzas
Code_upm	Identificador de unidad primaria
Est_urb	Estrato de urbanidad
Est_marg	Estrato de marginalidad
Est_dis	Estrato de diseño marginalidad x urbanidad
pondeh	Ponderador de hogar
pondei	Ponderador de integrantes

Importancia de los ponderadores

- Gráfica 1:
 - graph hbar (sum) hombres mujeres [pw=pondei], over (gpo_edad, descending gap(0)) stack
- Gráfica 2:
 - graph hbar (sum) hombres mujeres, over (gpo_edad, descending gap(0)) stack

Gráfica 1. Sin ponderadores



Gráfica 2. Con ponderadores

Unión de bases

Supongamos los siguientes conjuntos hipotéticos de datos

Folio	Datos Adulto-Salud
1	Si
2	Si
3	Si
4	Si

Folio	Datos Adulto-Antropometria
2	Si
3	Si
6	Si

Entonces, existen dos opciones de análisis .

- Copiar los datos de adulto-salud al archivo de adultos-antropometría
- Copiar los datos de adulto-antropometría al archivo de adultos-salud

Cual de las dos opciones es válida y cual es la mejor ?

Unión de bases

Las dos opciones son válidas.

Salud → Antropometría

Folio	Adultos -Salud	Adultos- Antropometría
1	Si	No
2	Si	Si
3	Si	Si
4	Si	No

Antropometría → Salud

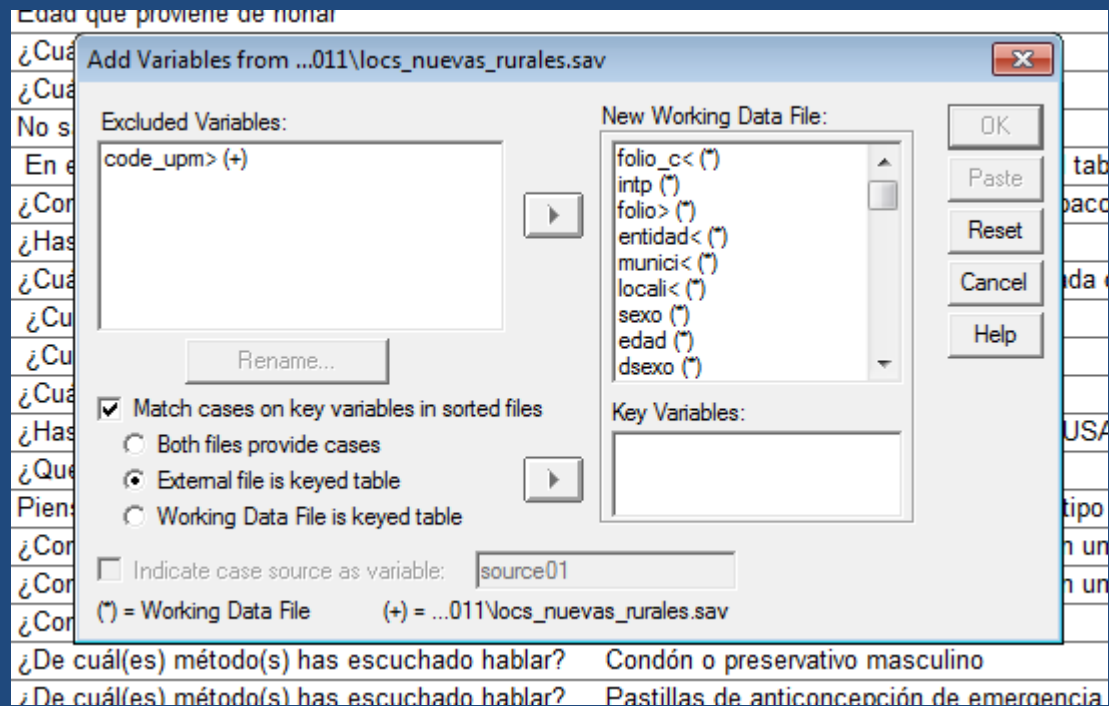
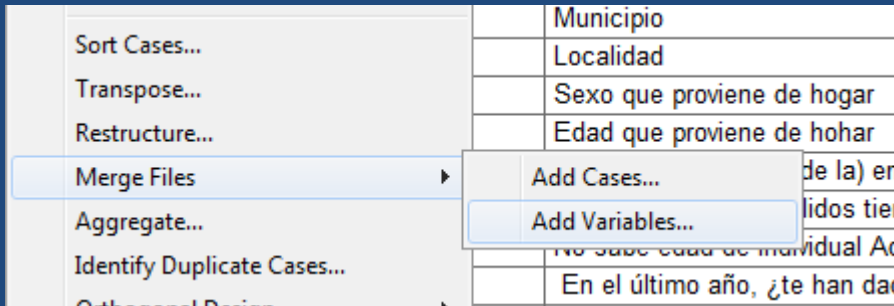
Folio	Adultos -Salud	Adultos- Antropometría
2	Si	Si
3	Si	Si
6	No	Si

La elección depende de cual de las dos bases de datos tiene menos valores faltantes

Llaves para pegar las bases

- La variable única por hogar
 - folio_c
- Para análisis por individuo, la variable que identifica al individuo
 - (folio_c, intp)
- En Stata
- `merge 1:1 folio_c intp using "dir\base.dta"`

Llaves para pegar bases



Gracias por su atención

Con gusto, cualquier duda, favor de escribir a info_ciee@insp.mx

