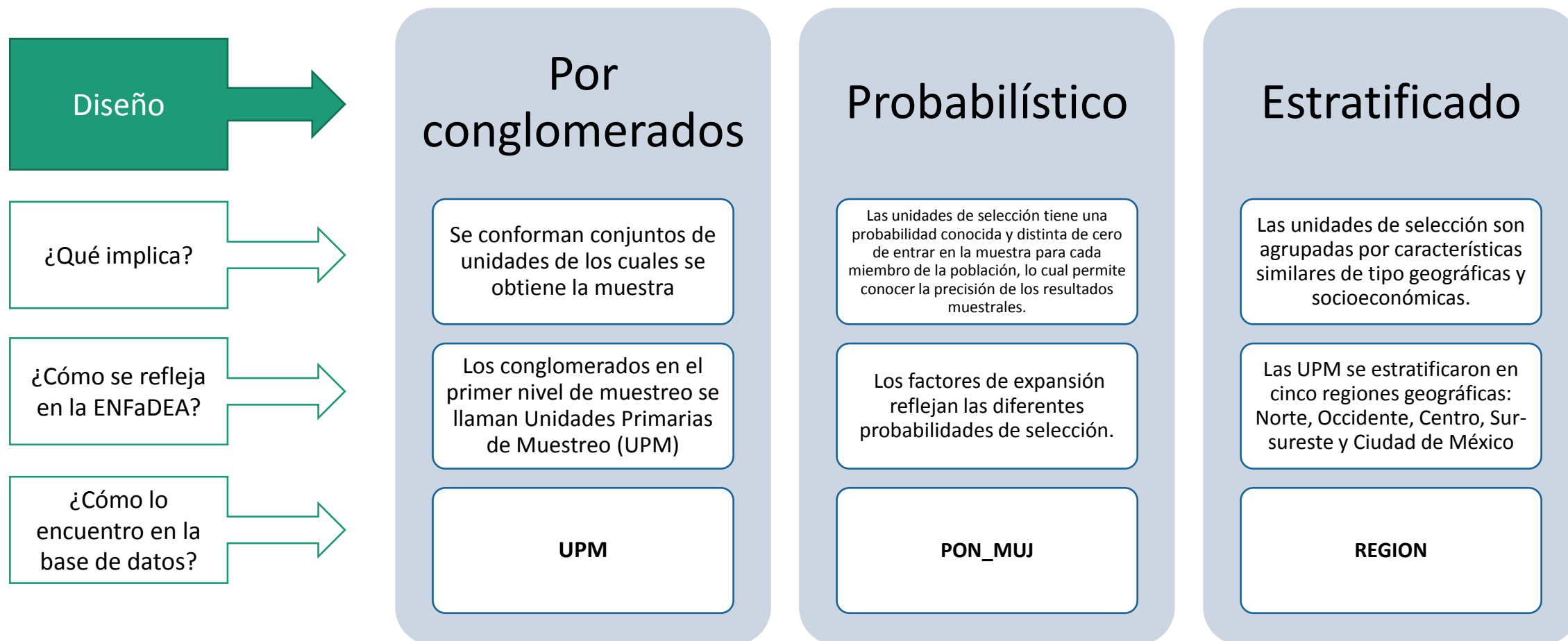




Guía para el uso del diseño complejo de la encuesta ENFaDEA 2017

Mariana Lugo Arellano
Escuela Nacional de Trabajo Social
UNAM
Mayo de 2018

Los elementos que indican el diseño complejo de la encuesta:



Declarar el diseño complejo de la ENFaDEA en Stata:

1. Se indican los elementos (**UPM**, **PON_MUJ** y **REGION**) usando el comando **svyset**:

- Forma:

svyset **upm** [**factor**], strata(**estrato**) opciones_extra

- Comando con los datos de la ENFaDEA:

svyset **UPM** [**pw= PON_MUJ**], strata(**REGION**) singleunit(certainty)

Nota: Hace que los estratos con unidades de muestreo individuales sean tratados como unidades de certeza

```
. svyset UPM [pw= PON_MUJ], strata(REGION) singleunit(certainty)

      pweight: PON_MUJ
      VCE: linearized
Single unit: certainty
Strata 1: REGION
  SU 1: UPM
  FPC 1: <zero>

. svydescribe
```

Survey: Describing stage 1 sampling units

```
      pweight: PON_MUJ
      VCE: linearized
Single unit: certainty
Strata 1: REGION
  SU 1: UPM
  FPC 1: <zero>
```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	3	795	191	265.0	342
2	3	683	56	227.7	398
3	3	671	68	223.7	442
4	2	286	121	143.0	165
5	2	945	78	472.5	867
5	13	3,380	56	260.0	867

En total tenemos 5 estratos, 13 UPM y 3380 observaciones

Una vez que se indica el diseño complejo de la encuesta con **svyset**:

- Usamos **svy** para ejecutar un comando y especificar el diseño complejo de la encuesta

svy: comando

Nota: Si desean ejecutar algún comando sin considerar el diseño complejo de la encuesta, simplemente se omite la indicación **svy**:

- Ejemplos:
 - *svy: tab variable*
 - *svy: logit variable variable*



- Para los descriptivos también es posible obtener estimaciones puntuales utilizando el factor de expansión (**PON_MUJ**)

Entre los comandos que se pueden ejecutar con el prefijo **svy** están los siguientes:

Descriptive statistics

mean	[R] mean — Estimate means
proportion	[R] proportion — Estimate proportions
ratio	[R] ratio — Estimate ratios
total	[R] total — Estimate totals

Survival-data regression models

stcox	[ST] stcox — Cox proportional hazards model
streg	[ST] streg — Parametric survival models

Binary-response regression models

biprobit	[R] biprobit — Bivariate probit regression
cloglog	[R] cloglog — Complementary log-log regression
hetprobit	[R] hetprobit — Heteroskedastic probit model
logistic	[R] logistic — Logistic regression, reporting odds ratios
logit	[R] logit — Logistic regression, reporting coefficients
probit	[R] probit — Probit regression
scobit	[R] scobit — Skewed logistic regression

Discrete-response regression models

clogit	[R] clogit — Conditional (fixed-effects) logistic regression
mlogit	[R] mlogit — Multinomial (polytomous) logistic regression
mprobit	[R] mprobit — Multinomial probit regression
ologit	[R] ologit — Ordered logistic regression
oprobit	[R] oprobit — Ordered probit regression
slogit	[R] slogit — Stereotype logistic regression

Para obtener la lista completa consultar: <https://www.stata.com/manuals13/svy.pdf>

Ejemplos para análisis descriptivo en Stata:

Sin factor de expansión

```
. tab ea
```

Embarazo adolescente	Freq.	Percent	Cum.
No	1,961	58.02	58.02
Sí	1,419	41.98	100.00
Total	3,380	100.00	



Indicando el diseño complejo de la encuesta

```
. svy: tab ea  
(running tabulate on estimation sample)
```

Number of strata	=	5	Number of obs	=	3,380
Number of PSUs	=	13	Population size	=	4,792,056
			Design df	=	8

Embarazo adolescente	proportion
No	.6188
Sí	.3812
Total	1



Key: proportion = cell proportion

Con factor de expansión

```
. tab ea [iw=PON_MUJ]
```

Embarazo adolescente	Freq.	Percent	Cum.
No	2,965,345	61.88	61.88
Sí	1,826,711.1	38.12	100.00
Total	4,792,056	100.00	



Con factor de expansión y con el diseño complejo de la encuesta se tienen los mejores estimadores

Con factor de expansión:

```
. tab AMBITO ea [iw=PON_MUJ], col
```

Key
frequency
column percentage

Ambito	Embarazo adolescente		Total
	No	Si	
Urbana	2295440.4 77.41	1252499.6 68.57	3,547,940 74.04
Rural	669,904.5 22.59	574,211.5 31.43	1,244,116 25.96
Total	2965344.9 100.00	1826711.1 100.00	4,792,056 100.00

En ambos casos se tienen los mejores estimadores

```
. tab AMBITO ea [iw=PON_MUJ], col chi2
option chi2 not allowed
r(198);
```

En Stata, no se puede realizar la prueba de χ^2 con el ponderador

Indicando el diseño complejo de la encuesta:

```
. svy: tab AMBITO ea, col
(running tabulate on estimation sample)
```

Number of strata = 5
Number of PSUs = 13

Number of obs = 3,380
Population size = 4,792,056
Design df = 8

Ambito	Embarazo adolescente		Total
	No	Si	
Urbana	.7741	.6857	.7404
Rural	.2259	.3143	.2596
Total	1	1	1

Key: column proportion

Pearson:

Uncorrected chi2(1) = 32.4362
Design-based F(1, 8) = 11.7354 P = 0.0090

Cuando se ejecuta el comando `svy: tab var1 var2` automáticamente se obtiene una prueba de χ^2 considerando el diseño complejo de la encuesta

Sin factor de expansión:



Se controla por el tamaño de la muestra, pero asume que es un muestreo aleatorio simple

```
. logit ea AMBITO dem_v05
```

```
Iteration 0:  log likelihood = -2299.1931
Iteration 1:  log likelihood = -2275.574
Iteration 2:  log likelihood = -2275.5589
Iteration 3:  log likelihood = -2275.5589
```

```
Logistic regression              Number of obs   =      3,380
                                LR chi2(2)        =       47.27
                                Prob > chi2         =      0.0000
                                Pseudo R2          =      0.0103

Log likelihood = -2275.5589
```

ea	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
AMBITO	.4939257	.0726292	6.80	0.000	.3515751	.6362764
dem_v05	.0087028	.0211884	0.41	0.681	-.0328257	.0502312
_cons	-1.140171	.1302431	-8.75	0.000	-1.395443	-.8848995

Con factor de expansión:



Se obtienen mejores estimadores, pero al expandir la muestra se exagera la significancia estadística.

```
. logit ea AMBITO dem_v05 [iw=PON_MUJ]
```

```
Iteration 0:  log likelihood = -3185023
Iteration 1:  log likelihood = -3161118.9
Iteration 2:  log likelihood = -3160188.6
Iteration 3:  log likelihood = -3160184.7
Iteration 4:  log likelihood = -3160184.7
```

```
Logistic regression              Number of obs   =      3,380
                                LR chi2(2)        =     49676.49
                                Prob > chi2         =      0.0000
                                Pseudo R2          =      0.0078

Log likelihood = -3160184.7
```

ea	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
AMBITO	.4549177	.0021154	215.05	0.000	.4507716	.4590637
dem_v05	-.124875	.0020769	-60.13	0.000	-.1289457	-.1208044
_cons	-.8044227	.0050863	-158.15	0.000	-.8143917	-.7944536

Ejemplo para análisis inferencial en Stata:

Indicando el diseño complejo de la encuesta:

```
. svy: logit ea AMBITO dem_v05
(running logit on estimation sample)
```



Survey: Logistic regression

```
Number of strata   =      5
Number of PSUs     =     13

Number of obs      =      3,380
Population size    =    4,792,056
Design df          =      8
F( 2, 7)           =      5.22
Prob > F           =     0.0410
```

ea	Linearized			P> t	[95% Conf. Interval]	
	Coef.	Std. Err.	t			
AMBITO	.4549177	.1322816	3.44	0.009	.1498757	.7599596
dem_v05	-.124875	.3194205	-0.39	0.706	-.86146	.6117099
_cons	-.8044227	.6906019	-1.16	0.278	-2.396954	.7881083



Permite obtener los mejores estimadores sin expandir la muestra. Los coeficientes tienen el mejor ajuste, se considera el diseño complejo de la encuesta y no se exagera el nivel de significancia

Resumiendo

- La **ENFaDEA** tiene un diseño complejo; tomar en cuenta sus características permite obtener las mejores estimaciones.
- Para indicar el diseño complejo de la encuesta en **STATA**:
 1. Se utiliza el comando **svyset** para señalar los parámetros:

```
svyset UPM [pw= PON_MUJ], strata(REGION)  
singleunit(certainty)
```
 2. El resto de los comandos se ejecutan indicando el diseño complejo de la encuesta con **svy**:

```
svy: tab [variable] ; svy: logit [variables] ; etc.
```
- Para el **análisis descriptivo** se puede utilizar el **factor de expansión** o el diseño complejo de la encuesta (**svy**); para el **análisis inferencial** se recomienda usar el diseño complejo de la encuesta (**svy**).



Algunas notas:

- El uso de restricciones como “if” podría no arrojar estimaciones de varianza correctas para las subpoblaciones. Por lo que se recomienda utilizar la opción **subpop ()**

- **Ejemplo:**

```
. svy, subpop (if AMBITO==1): tab ea  
(running tabulate on estimation sample)
```

```
Number of strata   =          5  
Number of PSUs    =          13
```

```
Number of obs      =        3,380  
Population size    =    4,792,056  
Subpop. no. obs    =        1,331  
Subpop. size       =    3,547,940  
Design df          =              8
```

Embarazo adolescente	proportion
No	.647
Si	.353
Total	1

Key: proportion = cell proportion



Algunas notas (*continuación*):

- El uso del diseño complejo de la encuesta también es recomendable para el tratamiento de **muestras pequeñas**, como es el caso del embarazo infantil.
- Algunas pruebas de postestimación no son válidas cuando se usa el diseño complejo de la encuesta. Ejemplo: estat class y estat gof

Ventajas de utilizar el diseño complejo de la encuesta en el análisis descriptivo e inferencial:

1

Considera las características probabilísticas, estratificadas y por conglomerados de la encuesta

2

Permite obtener estimaciones puntuales y los errores estándar correctos

3

Permite obtener los mejores estimadores sin exagerar la significancia estadística.

Para mayor información se recomienda consultar los siguientes manuales:

- STATA:

- <https://www.stata.com/manuals13/svy.pdf>

- SPSS:

- [http://www.sussex.ac.uk/its/pdfs/SPSS Complex Samples 22.pdf](http://www.sussex.ac.uk/its/pdfs/SPSS_Complex_Samples_22.pdf)